HARVARD UNIVERSITY
DEPARTMENT OF STATISTICS

Resampling Methods for Time Series

Ernesto Ramos
BBN Software Products Corp.
10 Fawcett St.
Cambridge, MA 02238

Technical Report No. ONR-C-2
August, 1988

Contents

Abstract

Suppose we observe a realization of size n of a Gaussian stationary sequence and we estimate θ_f, a functional of the spectral density, by $\hat{\theta}_x$. For example, θ_f and $\hat{\theta}_x$ might be the true and the empirical rate at which the process crosses a certain level. It is natural to ask: (i) can the sampling distribution of $\hat{\theta}_x$ be consistently estimated? and (ii), can a better estimator of θ_f than $\hat{\theta}_x$ be constructed on the basis of $\hat{\theta}_x$?

In this paper we describe approaches which allow answering both questions affirmatively in certain situations. The approach is based on *resampling* the data, i.e., using the original data to specify a distribution from which new samples are drawn. In the first problem the data induce an estimate of the spectral density $\hat{f}_n$ and the sampling distribution of $\hat{\theta}_x$ under $\hat{f}_n$ is used to estimate that under f. We establish results for estimators of well-behaved functions of linear functionals of the spectral density. In the second problem we pretend the data are circular so that the periodogram of the data is a sufficient statistic and thus the conditional expectation of $\hat{\theta}_x$ is a better estimator. If the data are not really circular, this mechanism introduces bias but the reduction in variance may be substantial enough to reduce the mean square error.

Chapter 1

Introduction

1.1 Setup and problems

Suppose we observe $X_0^{n-1} \sim N(0, \Sigma_n)$, a realization of size n of a Gaussian stationary process $\{X_t\}_{-\infty}^{\infty}$. Since the process is stationary Σ_n is a Toeplitz matrix and completely specified for all n by a spectrum f, and we may write $X_0^{n-1} \sim N(0, f)$. Furthermore, suppose we estimate $\theta_f = \theta(f)$, some functional of f, by procedure $\hat{\theta}_x$.

In this paper we explore the use of resampling methods in connection with two problems:

1. consistent estimation of the sampling distribution of $\hat{\theta}_x$ and
2. using $\hat{\theta}_x$ to find a better estimator of θ_f.

1.2 General strategy

The basic strategy behind our approach to each of these problems is based on *resampling* the observed data, i.e., using the observed data X_0^{n-1} to specify a distribution from which we then draw samples Y_0^{n-1} which are sometimes called *bootstrap* samples.

There are two steps implicit in this approach: the first defines the procedure through which the original data specify the distribution from which new samples are to be drawn; the second indicates how these bootstrap samples are to be used to solve each of the two problems at hand, e.g., using Monte-Carlo to approximate distributions and their moments.

1.2.1 Strategy for problem 1

For the first problem we proceed as follows:

1. Use X_0^{n-1} to estimate the spectrum f by $\hat{f}_n$ via some procedure.
2. Draw $Y_0^{n-1} \sim N(0, \hat{f}_n)$ and calculate $\hat{\theta}_y$ for each sample
3. Take the Monte-Carlo approximation of the (conditional on X_0^{n-1}) sampling distribution of $\hat{\theta}_y$ as the estimate of the sampling distribution of $\hat{\theta}_x$.

1.2.2 Strategy for problem 2

For the second problem we proceed as follows:

1. Use X_0^{n-1} to construct a sufficient statistic $S_n = S(X_0^{n-1})$.

2. Draw samples $Y_0^{n-1} \sim \mathcal{L}(X_0^{n-1}|S_n)$ and calculate θ_y for each

3. Approximate $\hat{\theta}_x^* = E(\hat{\theta}_x|S_n)$ by the average of the $\hat{\theta}_y$'s.

1.3 Results for problem 1

In order to be able to give fairly exhaustive coverage to problem 1 we will restrict attention to estimates of linear functionals of the spectral density and well-behaved functions of these.

1.3.1 Linear functionals

Linear functionals of a spectrum f are of the form

$$\theta_f = \int \theta(\omega) f(\omega) d\omega$$

and they can be efficiently estimated (cf., e.g., Hasminskii and Ibragimov, 1986) by the corresponding linear functional of the periodogram I_n^x of the data

$$\hat{\theta}_x = \int \theta(\omega) I_n^x(\omega) d\omega.$$

For example, covariances and their MLE's (under the Toeplitz model) are of this form.

1.3.2 Results for linear functionals

It simplifies matters to estimate the sampling distribution of $U_n = \sqrt{n}(\hat{\theta}_x - \theta_f)$ rather than that of $\hat{\theta}_x$. As mentioned above, we first use X_0^{n-1} to estimate the spectrum f by $\hat{f}_n$ and then draw samples Y_0^{n-1} from $\mathcal{L}(Y_0^{n-1}|X_0^{n-1}) = N(0, \hat{f}_n)$. If we let $V_n = \sqrt{n}(\hat{\theta}_y - \theta_{\hat{f}_n})$, we can then use $\mathcal{L}(V_n|X_0^{n-1})$ to estimate $\mathcal{L}(U_n)$.

The device used in deriving our results is to construct $\tilde{U}_n$ and $\tilde{V}_n$ on the same probability space with the property that $\mathcal{L}(\tilde{U}_n) = \mathcal{L}(U_n)$ and $\mathcal{L}(\tilde{V}_n|X_0^{n-1}) = \mathcal{L}(V_n|X_0^{n-1})$.

Our main result, Corollary 3.5, establishes that, under regularity conditions on $\hat{\theta}$, f, and $\hat{\theta}_n$,

$$d_P(\mathcal{L}(U_n), \mathcal{L}(\tilde{V}_n|X_0^{n-1})) = o_p(1),$$

i.e., the Prohorov distance between the distribution of interest and its estimate is small with high probability.

The proof of this result is an easy consequence of Theorem 3.4 which establishes $\tilde{U}_n - \tilde{V}_n = o_p(1)$ under the following regularity conditions:

1. f, θ satisfy a Lipschitz condition.

2. $\|f - \hat{f}_n\|_2^2 = \int_0^1 |f(\omega) - \hat{f}_n(\omega)|^2 d\omega = o_p(1)$

3. $P(\inf_\omega \hat{f}_n(\omega) \geq 0) \to 1$

4. $\sup_\omega \hat{f}_n(\omega) = O_p(1)$

5. $\sup_\omega \hat{f}_n^{-1}(\omega) = O_p(1)$

6. $\sup_\omega |\hat{f}_n'(\omega)|^2 = o_p(n/\log^2 n)$.

We then extend these results to differentiable functions h of linear functionals in Theorem 3.8 and Corollary 3.9, in which case the variables of interest are

$$U_n^* = \sqrt{n}\left(h(\hat{\theta}_{x1}, \ldots, \hat{\theta}_{xp}) - h(\theta_{f1}, \ldots, \theta_{fp})\right)$$

and

$$V_n^* = \sqrt{n}\left(h(\hat{\theta}_{y1}, \ldots, \hat{\theta}_{yp}) - h(\theta_{f_n1}, \ldots, \theta_{f_np})\right).$$

1.3.3 Spectrum estimation

The motivating idea for assessing sampling distributions is a practical one: we would like to use this method in real-life applications, for example, to estimate the sampling bias and variability of statistics of interest. We mentioned in the previous section that this resampling strategy will only work if the estimate of the spectral density used to generate the bootstrap samples satisfies certain properties.

In our simulations we implemented the spectrum estimation technique derived in Wahba (1980) which estimates the logarithm of the spectral density by smoothing the log of the periodogram.

We chose Wahba's procedure for two reasons: (i) there is an easily programmable objective criterion for choosing the amount of smoothing which makes the procedure automatic and thus suitable for large simulation studies; (ii) the degree of differentiability of the estimate can be controlled.

The amount of smoothing and the degree of differentiability of Wahba's estimate are controlled by two tuning parameters. Choosing the value of the smoothing parameters is essentially a choice between controlling bias or controlling variability (and hence regularity properties) of the resulting estimate of the spectral density.

Specifically, let λ and m denote the two tuning parameters: λ controls the amount of smoothing and m represents the number of continuous derivatives of the estimate. Let h be the log of the spectral density and $\hat{h}_n$ be its estimate so that $\hat{f}_n = \exp(\hat{h}_n)$ is the estimate of f. Then, by letting λ be a function of the sample size n, i.e., picking a sequence $\lambda_n = \lambda(n)$, which decays to 0 slowly enough, certain regularity properties of $\hat{h}_n$ (and hence of $\hat{f}_n$) can be controlled. In particular, we establish that if $\lambda_n = O(n^{-r})$ and $0 < r < m/3$, then

1. $\sup_\omega |\hat{h}_n(\omega)| = O_p(1)$ which implies
   (a) $\sup_\omega \hat{f}_n(\omega) = O_p(1)$ and
   (b) $\sup_\omega \hat{f}_n^{-1}(\omega) = O_p(1)$
2. $\sup |h_n'(\omega)|^2 = O_p(n^{1/3})$ which, together with condtion (1.a), implies $\sup |f_n'(\omega)|^2 = O_p(n^{1/3})$.
3. $E \int |h(\omega) - \hat{h}_n(\omega)|^2 d\omega = o(1)$ which, together with condition (1.a), implies $\int |f(\omega) - \hat{f}_n(\omega)|^2 d\omega = o_p(1)$.

These properties can be seen to be sufficient for the conditions described at the end of the previous section to hold, and thus enable the results of this paper to obtain in practical applications where the parameter of interest can be expressed as a well-behaved function of linear functionals of the spectral density.

1.3.4 Applications

We apply this resampling methodology to the problem of estimating the first correlation ρ of the process by the sample correlation $\hat{\rho} = \sum X_t X_{t+1} / \sum X_t^2$. Since ρ and $\hat{\rho}$ can be expressed as ratios of linear functionals of the spectral density and periodogram respectively, the theoretical results obtained apply in this case.

The simulations studies indicate that the estimate of variability obtained by our method are roughly as good as the asymptotic approximations, and much better when estimating bias, even though the asymptotics were calculated assuming complete knowledge of the model generating the data.

1.4 Improved estimators

Suppose we observe a realization of size n, X_0^{n-1}, of a Gaussian stationary process with spectral density f, and we estimate a general functional $\theta_f = \theta(f)$ by $\hat{\theta}_x$.

1.4.1 Circular process

Under the assumption that the process is circular the periodogram I_n^x is a sufficient statistic and we can consider the estimator $\hat{\theta}_x^* = E\left\{\hat{\theta}_x | I_n^x\right\}$ which has smaller risk than $\hat{\theta}_x$ under convex loss by the Rao-Blackwell Theorem.

Under the assumption of circularity it is easy to calculate $\hat{\theta}_x^*$ since, conditional on the periodogram, the first half of the discrete Fourier transform of the data has fixed amplitude and independent uniform $U(0,1)$ phases. Thus, by sampling iid $U(0,1)$ phases we can obtain new samples of the discrete Fourier transform (DFT) of the original data with the same conditional (given the periodogram) distribution. Taking inverse DFT we can obtain new samples of data Y_0^{n-1}, calculate $\hat{\theta}_y$ for each, and average to obtain an approximation to $\hat{\theta}_x^*$.

1.4.2 Non-circular process

If the original process is not circular the periodogram is not a sufficient statistic but we may pretend it is and resample as in the circular case to obtain $\hat{\theta}_x^*$. However, this mechanism will in general introduce bias, i.e., $E\hat{\theta}_x^* \neq E\hat{\theta}_x$, but the amount of bias is often neglibible compared to the reduction in variance so that $\hat{\theta}_x^*$ has lower mean square error than $\hat{\theta}_x$.

1.4.3 Applications

The motivating idea is to obtain more efficient estimators in situations where efficient estimators, e.g., MLE, cannot be calculated easily but there exist inefficient estimators. This happens, for example, in the context of estimating the crossing rate of level α of a Gaussian stationary process. In this case the MLE (undet the Toeplits model) can only be evaluated in the case $\alpha = 0$ in which case it is given by $.5 - \pi^{-1} \arcsin(\hat{\rho})$, where $\hat{\rho}$ is the first sample correlation coefficient. On the other hand, the empirical crossing rate $\hat{\phi}_\alpha^o$ which simply counts the number of crossings and divides by $n-1$, can be calculated for all α.

The simulations indicate that, for the case $\alpha = 0$, the MSE of $E\left\{\hat{\phi}_\alpha^o | I_n^x\right\}$ is roughly equal to that of the MLE, and much smaller than that of $\hat{\phi}_\alpha^o$. Furthermore, it is worth mentioning again that our procedure has the the advantage over the MLE that it can be calculated for all α while the MLE is limited to the case $\alpha = 0$.

1.5 Algorithms

Resampling plans are by nature computer intensive. For this reason, one of the emphases of this research was to implement algorithms which could be driven by the FFT. To generate n-long realizations of stationary Gaussian processes with covariance matrix Σ_n we adapted an algorithm to simulate circular processes of length $2n$ due to Jonas (1984) whose n-long segments have covariance Σ_n.

1.6 Related research

The idea of using the data to specify a distribution from which to resample was advanced, in the iid case, by Efron (1979). The idea of resampling Gaussian stationary process with spectral density specified by the data to study sampling bias and variability was set forth in Ramos (1984). Subsequently Hurvich and Zeger (1986) and Stine (1986) have come up with similar schemes for different applications. In related research, methods to estimate variability in the context of time-dependent observations have been suggested by Carlstein (1986) who estimates it on the basis of subseries values and by Freedman and Peters (1984) who, in the original spirit of Efron (1979), suggest fitting a long autoregression and generating new psuedo-series by resampling the residuals.

1.7 Contents

The paper is organized in the following manner:

1.7.1 Chapter 2

In this chapter the background and numerous auxiliary results are presented. Of particular interest are results concerning the magnitude of the covariance between periodogram ordinates and their logs, Lemmas 2.16 and 2.17.

1.7.2 Chapter 3

In this chapter we establish the results concerning linear functionals. The relevant results are Corollaries 3.5 and 3.9 which establish tha. the bootstrap distribution is close, in probability, to the unknown sampling distribution of the estimator.

1.7.3 Chapter 4

In this chapter we establish the properties of Wahba's (1980) procedure. The main results are Lemmas 4.3, 4.4, 4.8 and 4.7 which establish regularity properties and convergence in L_2 norm to 0 of the integrated mean square error.

1.7.4 Chapter 5

In this chapter we describe the Monte-Carlo algorithm for simulating realizations from Gaussian stationary processes.

1.7.5 Chapter 6

In this chapter we describe applications of our resampling methodolgy to the problem of estimating the sampling bias and variability of the first sample correlation under the AR(1) and MA(1) models and result of simulations.

1.7.6 Chapter 7

In this chapter we examine the problem of finding better estimators by resampling. An application to estimating the crossing rate of a certain level is discussed and results of simulations are presented.

1.7.7 Appendix A

All proof of results are given in the Appendix.

Chapter 2

Background

In this chapter we state the basic definitions, assumptions, and properties of stationary Gaussian processes and their spectra; we also define the discrete Fourier transform (DFT) and the periodogram, investigate their distributions and moments, and review some asymptotics. Finally, we briefly discuss circulants with two examples which will be useful later on.

2.1 Gaussian stationary processes

In this section we introduce Gaussian stationary process and investigate properties of their covariance structure and their spectra. We also describe a representation of such processes due to Cramer.

2.1.1 Covariance structure

In what follows let $\{X_t\}_{t=-\infty}^{\infty}$ (or more simply $\{X_t\}$) denote a real and discrete time series, that is, a random sequence $\ldots, X_{-1}, X_0, X_1, \ldots$. In this paper we will only be concerned with time series which are stationary and Gaussian. Intuitively speaking, stationarity means that the distribution of the X's does not change with time; formally:

Definition 2.1 *$\{X_t\}$ is a Gaussian stationary sequence if all its finite dimensional distributions are multivariate normal, $E(X_t) = \mu$, and $\mathrm{cov}(X_t X_{t+s}) = c_s$, for all t, s.*

Remark: For convenience and without undue loss of generality, we will assume throughout this paper that $\mu = 0$.

Stationarity implies that if $X_0^{n-1} = (X_0, \ldots, X_{n-1})'$ is a realization of size n of a stationary Gaussian process with covariance sequence $\{c_s\}$, its covariance matrix Σ_n is *Toeplitz* with first row $c_0, \ldots, c_{n-1}$. Recall that $\Sigma_n = (\sigma_{jk})$ is Toeplitz whenever $\sigma_{jk} = \sigma_{|j-k|}$, i.e., the value of any entry is a function only of its distance to the main diagonal. Thus, it is easy to see that the first row (or column) defines any Toeplitz Σ_n and thus it is convenient to write $\Sigma_n = T(c_0, \ldots, c_{n-1})$ where T maps the column vector $(c_0, \ldots, c_{n-1})'$ to the Toeplitz matrix with first row $(c_0, \ldots, c_{n-1})$. An easy consequence of definition 2.1 is that, for all n and s, $X_s^{n-1+s} \sim N(0, \Sigma_n)$ where $\Sigma_n = T(c_0, \ldots, c_{n-1})$.

2.1.2 Spectrum

A dual representation of the covariance structure of a stationary process is given by the spectrum.

Definition 2.2 *If* $\sum_{-\infty}^{\infty} |c_s| < \infty$ *the spectrum of the process can be defined by*

$$f(\lambda) = \sum_{s=-\infty}^{\infty} c_s \exp(i2\pi s\lambda), \tag{2.1}$$

for all λ.

Assumption 2.1 *The covariances of all processes considered in this paper satisfy* $\sum_{-\infty}^{\infty} |c_s| < \infty$.

The characteristic properties of the spectrum are:

1. f is clearly periodic with period 1 so we can limit the range of λ to any interval of unit length.
2. Since $\{X_t\}$ is real-valued, $c_s = c_{-s}$ and hence $f(\lambda) = f(-\lambda)$ which, together with the fact that f is periodic, implies f is symmetric about $1/2$, i.e., $f(1/2 - x) = f(1/2 + x)$.
3. f and $\{c_s\}$ are a Fourier transform pair and hence

$$c_s = \int_a^{a+1} f(\lambda) \exp(-i2\pi s\lambda)\, d\lambda. \tag{2.2}$$

   for all real a and integer s.

4. *Brillinger (1980, p. 24) shows that Definition 2.2 implies* $f \geq 0$.

Lemma 2.1 *If* $f > 0$, *then* $T(c_0, \ldots, c_{n-1}) > 0$ *for all* n.

Remark: The spectrum determines $T(c_0, \ldots, c_{n-1})$ for all n so that we will write $X_0^{n-1} \sim N(0, f)$ to indicate $X_0^{n-1} \sim N(0, T(c_0, \ldots, c_{n-1}))$.

Smoothness of the spectrum

It is easily seen from (2.2) that a constant spectrum implies $c_s = 0$ for $s \neq 0$, i.e., the $\{X_t\}$ are independent and the process is memoryless. On the other hand, if the spectrum has a sharp peak near the origin the c_s's will decay slowly and we can think of the process as having a long memory.

In the next few paragraphs we briefly discuss in what way "smoothness" properties of the spectrum and conditions on the covariance sequence $\{c_s\}$ are related and state the "smoothness" conditions we shall impose on the spectra under consideration in this paper.

"Smoothness" of a function f is usually quantified in one of two ways: (i) by properties of f, namely Lipschitz conditions and differentiability, or (ii) properties of the Fourier coefficients $\{c_s\}$ of f, namely the rate of decay to 0 of $\{c_s\}$.

Recall the definition of a Lipschitz condition:

Definition 2.3 $f \in \mathrm{Lip}(\alpha)$ *whenever there exists a smallest constant* $M_{f,\alpha}$ *such that*

$$\sup_{\omega, |h| \leq \epsilon} |f(\omega + h) - f(\omega)| \leq M_{f,\alpha} \epsilon^{\alpha}$$

for all $\epsilon > 0$.

Remarks:

1. $M_{f,\alpha}$ is called the *α-modulus of continuity* of f.

2. We will write M_f for $M_{f,1}$.

It is clear from definition 2.2 that $\sum |c_s| < \infty$ *implies* $\sup_x f(x) < \infty$, although the converse does not necessarily hold. In general, conditions on how quickly the sequence $\{c_s\}$ converges to zero, differentiability of f, and Lipschitz conditions are related in the following manner:

Lemma 2.2 *If* $f(\omega) = \sum_{s=-\infty}^{\infty} c_s \exp(i2\pi s\omega)$ *then*

$$\sum |s|^k |c_s| < \infty \Longrightarrow \sup_\omega |f^{(k)}(\omega)| < \infty \Longrightarrow f^{(k-1)} \in \text{Lip}(1)$$

for $k \geq 1$, *and Assumption 2.1 implies that* f *is continous.*

Assumption 2.2 *In this paper we shall only consider processes for which the spectrum* $f \in$ Lip(1) *and* $f > 0$.

2.1.3 Cramer representation

Throughout this paper we will appeal to a representation of Gaussian stationary processes due to Cramer. Its importance lies in the fact that it expresses the values of the process as a simple function of the spectrum and is thus useful in establishing results which depend on properties of the spectrum.

Theorem 2.3 (Cramer) *If* $\{X_t\} \sim N(0, f)$ *we can write*

$$X_t = \int_{-1/2}^{1/2} \exp(i2\pi t\omega) F(\omega) dZ(\omega) \tag{2.3}$$

where F *is any integrable complex function such that* $F(\omega) = \overline{F(-\omega)}$ *and* $|F(\omega)|^2 = f(\omega)$, *and* $Z(\omega)$ *is a continuous and complex-valued process with independent Gaussian increments having the following properties:*

1. $EdZ(\omega) = 0$ *for all* ω,
2. $EdZ(\omega)\overline{dZ(\lambda)} = \delta_{\omega-\lambda} d\omega$,
3. $EdZ(\omega)dZ(\omega) = 0$,
4. $dZ(\omega) = \overline{dZ(-\omega)}$.

where δ_x *is Kronecker's* δ, *i.e.,* $\delta_x = 1$ *if* $x = 0$ *and* 0 *otherwise.*

Remarks:

1. The last property ensures the process is real-valued.
2. In the engineering literature F is called a *transfer function.*
3. If $f \in$ Lip(1) it will be convenient to take the transfer function $F \in$ Lip(1) too. In other words, let $F(\omega) = \sqrt{f(\omega)} \exp(i\theta(\omega))$ where $\theta \in$ Lip(1). For our purposes we rarely need to consider cases where $\theta(\omega) \neq 0$.

2.2 The main construction (MC)

In order to establish the main results of this paper it is necessary to construct two processes $\{X_t\} \sim N(0, f)$ and $\{Y_t\} \sim N(0, g)$ defined on the same probability space. These two processes are defined in such a way that they are to be "close" (in a sense to be made precise below) to each other, subject to the restriction $\{X_t\} \sim N(0, f)$ and $\{Y_t\} \sim N(0, g)$.

First we state the assumption on the transfer functions:

Assumption 2.3 *Let* $F, G \in$ Lip(1) *be measurable, complex-valued, non-vanishing functions such that* $F(\omega) = \overline{F(-\omega)}$, $|F(\omega)|^2 = f(\omega)$, $G(\omega) = \overline{G(-\omega)}$, *and* $|G(\omega)|^2 = g(\omega)$.

The main construction: Define two processes $\{X_t\}$ and $\{Y_t\}$ by

$$X_t = \int_{-1/2}^{1/2} \exp(i2\pi t\omega) F(\omega) dZ(\omega) \tag{2.4}$$

and

$$Y_t = \int_{-1/2}^{1/2} \exp(i2\pi t\omega) G(\omega) dZ(\omega) \tag{2.5}$$

for all t, where $Z(\omega)$ is as in Theorem 2.3, and F and G satisfy Assumption 2.3. We shall henceforth refer to this construction as the *Main Construction* (MC).

Notice that, under the conditions of the MC, for given spectra f and g the corresponding transfer functions F and G are defined up to amplitude but not phase so that different choices of phase functions will lead to processes which are more or less "close" to each other. If we measure "closeness" by $E|X_t - Y_t|^2$, it is then straightforward to show that

Lemma 2.4 *If $\{X_t\}$ and $\{Y_t\}$ are given by the MC, then*

$$E|X_t - Y_t|^2 = \int |F(\omega) - G(\omega)|^2 d\omega$$

for all t, and is minimized when the phase functions of F and G are the same, in which case $E|X_t - Y_t|^2 = \int |\sqrt{f(\omega)} - \sqrt{g(\omega)}|^2 d\omega$.

2.3 The DFT and the periodogram

In this section we define the discrete Fourier transform (DFT) and the periodogram associated with a realization X_0^{n-1}. We also examine their sampling properties when $X_0^{n-1} \sim N(0, f)$, $f \in \text{Lip}(1)$.

2.3.1 Definition of the DFT and periodogram

Definition 2.4 *Let $J_n^x(\omega) = n^{-1/2} \sum_{t=0}^{n-1} X_t e^{-i2\pi\omega t}$ denote the discrete Fourier transform of X_0^{n-1}.*

Definition 2.5 *Let $I_n^x(\omega) = |J_n^x(\omega)|^2$ be the associated periodogram.*

Remarks:

1. $J_n^x(\omega)$, $I_n^x(\omega)$ are periodic with period 1.
2. If X_0^{n-1} is real-valued it follows that $J_n^x(\omega) = \overline{J_n^x(1-\omega)}$. Hence $I_n^x(\omega) = I_n^x(1-\omega)$ and $I_n^x(1/2+\omega) = I_n^x(1/2-\omega)$.

We will write J_n^x without the argument ω to denote the column vector with entries $J_n^x(k/n)$, $k = 0, \ldots, n-1$, and similarly I_n^x is the vector whose components are $|J_n^x(k/n)|^2$, $k = 0, \ldots, n-1$. These are just the DFT and the periodogram evaluated at the set of *Fourier frequencies* k/n, $k = 0, \ldots, n-1$. We will refer to these vectors simply as the DFT and the periodogram.

2.3.2 Convolution and kernels

Since many of the results we derive below are expressed in terms of convolutions of a function f and a kernel, either Fejer's or Dirichlet's, we state the the relevant definitions and properties below. In this section f and g denote arbitrary periodic functions (not only spectral densities).

Definition 2.6 *If f and g are periodic define the convolution $f * g$ of f and g by*

$$f * g(x) = \int_a^{1+a} f(x-t)g(t)dt = \int_a^{1+a} f(t)g(x-t)dt.$$

Remarks:

1. We will usually take $a = 0$ or $a = -1/2$.
2. Notice that if either f of g are differentiable, their convolution is differentiable and

$$(f * g(x))' = \int_0^1 f'(x-t)g(t)dt = \int_0^1 f(t)g'(x-t)dt.$$

In this paper we will use the following notation: let L_p denote the space of periodic (almost everywhere) functions which are Lebesgue integrable to the p^{th} power with norm

$$||f||_p = \left\{ \int_{-1/2}^{1/2} |f(\omega)|^p d\omega \right\}^{1/p},$$

i.e., $f \in L_p$ if, and only if, $||f||_p < \infty$. If f and g are in L_p we will occasionally write $d_p(f,g) = ||f-g||_p$.

Lemma 2.5 *Let $f, g \in L_1$, $g \geq 0$, and $||g||_1 = 1$. Then*

$$\inf_\omega f(\omega) \leq \inf_\omega f * g(\omega) \leq \sup_\omega f * g(\omega) \leq \sup_\omega f(\omega).$$

In the following lemma S denotes either (i), the set of periodic continuous functions with sup norm, or (ii), L_p.

Lemma 2.6 *(Butzer and Nessel, 1971). Let $f \in S$ and $g \in L_1$. Then $f * g \in S$ and $||f * g||_S \leq ||f||_S ||g||_1$.*

Remark: If $||g||_1 = 1$ we can apply Lemma 2.6 to the derivatives of f to show that the result of convolving f and g is smoother than f.

We now state definitions and properties of the Dirichlet and the Fejer kernels.

Definition 2.7 *For $n \geq 0$ let*

$$D_n(\lambda) = \frac{\sin(n\pi\lambda)}{\sin(\pi\lambda)}$$

denote Dirichlet's kernel and let

$$K_n(\lambda) = \frac{\sin^2(n\pi\lambda)}{n\sin^2(\pi\lambda)} = \frac{1}{n} D_n^2(\lambda)$$

denote Fejer's Kernel.

These kernels arise naturally when approximating f by finite "sums" of its Fourier coefficients $\{c_k\}$. Here the word "sum" may refer to different types of sums, e.g. Cesaro. Each of these "sums" can be expressed as the convolution of f with a particular kernel.

The following are standard results from approximation theory (cf., e.g., Butzer and Nessel, 1971).

Lemma 2.7 *(Butzer and Nessel, 1971) If D_n and K_n are the Dirichlet and Fejer kernels respectively, then*

$$\sum_{k=-n}^{n} \exp(i2\pi k\lambda) = D_{2n+1}(\lambda), \tag{2.6}$$

$$\exp(-i\pi(n-1)\lambda) \sum_{k=0}^{n-1} \exp(i2\pi k\lambda) = D_n(\lambda) \tag{2.7}$$

and

$$\sum_{k=-n}^{n} \left(1 - \frac{|k|}{n}\right) \exp(i2\pi k\lambda) = K_n(\lambda). \tag{2.8}$$

Remark: Since $K_n \geq 0$, Lemma 2.7 implies $\|K_n\|_1 = 1$.

Lemma 2.8 *If $f \in L_1$ and $c_k = \int_0^1 f(\lambda)\exp(-i2\pi s\lambda)\,d\lambda$, then*

$$\sum_{k=-n}^{n} c_k \exp(i2\pi kx) = D_{2n+1} * f(x)$$

and

$$\sum_{k=-n}^{n} c_k \left(1 - \frac{|k|}{n}\right) \exp(i2\pi kx) = K_n * f(x).$$

Lemma 2.9 *(Butzer and Nessel, 1971) If $f \in \mathrm{Lip}(1)$, then*

$$\sup_{\omega} |K_n * f(\omega) - f(\omega)| = M_f O(\log n/n).$$

uniformly in f.

The next is an auxiliary lemma we will use often.

Lemma 2.10 *If D_n is the Dirichlet kernel, then*

$$\int_a^{a+1} D_n(x-t)D_n(x-s)dx = D_n(s-t).$$

for all a. If in addition $f \in Lip(1)$ is symmetric and periodic, then

$$\int_a^{a+1} \{f(x) - f(s)\}D_n(x-t)D_n(x-s)dx = M_f O(\log n).$$

uniformly in f and hence

$$\int_a^{a+1} f(x)D_n(x-t)D_n(x-s)dx = D_n(s-t)f(s) + M_f O(\log n).$$

uniformly in f for all s, t, and a.

2.3.3 Distribution and moments of the DFT

Let F_n denote the matrix with entries $n^{-1/2}\exp(-i2\pi jk/n)$, $j,k = 0,...,n-1$. F_n is called the Fourier transform matrix since it projects a vector onto its DFT, e.g., $J_n^x = F_n X_0^{n-1}$. Notice that F_n is symmetric and unitary, i.e. $F_n^{-1} = \overline{F_n}$. F_n should not be confused with F, the transfer function.

If we write $F_n = C_n - iS_n$ for real-valued matrices C_n and S_n, we can establish the following result concerning the distribution of J_n^x:

Lemma 2.11 *If $J_n^x = F_n X_0^{n-1}$ and $X_0^{n-1} \sim N(0,\Sigma_n)$, then*

$$\begin{pmatrix} \Re J_n^x \\ \Im J_n^x \end{pmatrix} \sim N\left(0, \begin{pmatrix} C_n\Sigma_n C_n & -C_n\Sigma_n S_n \\ -S_n\Sigma_n C_n & S_n\Sigma_n S_n \end{pmatrix}\right).$$

and $E|J_n^x(\omega)|^2 = EI_n^x(\omega) = K_n * f(\omega)$ for all ω.

The following lemma gives the variances and covariances between real and imaginary parts of the DFT.

Lemma 2.12 *If $X_0^{n-1} \sim N(0,f)$ for $f \in \text{Lip}(1)$ and $0 \le l,m \le [n/2]$, then*

$$2E\Re\{J_n^x(l/n)\}\Re\{J_n^x(m/n)\} = \delta_{m-l}(1+\delta_l+\delta_{n/2-l})K_n * f(m/n) + M_f O(\log n/n),$$

$$2E\Im\{J_n^x(l/n)\}\Im\{J_n^x(m/n)\} = \delta_{m-l}(1-\delta_l-\delta_{n/2-l})K_n * f(m/n) + M_f O(\log n/n),$$

and

$$E\Re\{J_n^x(l/n)\}\Im\{J_n^x(m/n)\} = (1-\delta_m)(1-\delta_{n-m/2})M_f O(\log n/n)$$

uniformly in $f \in Lip(1)$.

2.3.4 Distribution and moments of the periodogram

Distribution of periodogram ordinates

We saw above (Lemmas 2.11 and 2.12) that the real and imaginary parts of $J_n^x(k/n)$ have bivariate normal distribution with variance approximately proportional to $K_n * f(k/n)/2$ and correlation $\rho = M_f O(\log n/n)$. Lemma 2.11 implies that twice the periodogram ordinate $I_n^x(k/n)$ is distributed like $WK_n * f(k/n)$ where $W = X^2 + (\rho X + \sqrt{1-\rho^2}Y)^2$ and X and Y are iid $N(0,1)$ variates. The following lemma establishes an alternate representation for W which will make manipulations easier.

Lemma 2.13 *If $W = X^2 + (\rho X + \sqrt{1-\rho^2}Y)^2$ and X and Y are iid $N(0,1)$, then W is distributed as*

$$\begin{aligned} W &= (1+\rho)U_1^2 + (1-\rho)U_2^2 \\ &= (U_1^2+U_2^2)\left(1+\rho\frac{U_1^2-U_2^2}{U_1^2+U_2^2}\right) \\ &= (U_1^2+U_2^2)(1+\rho\theta) \end{aligned}$$

for U_1, U_2 iid $N(0,1)$. The two factors on the r.h.s. are independent and $|\theta| \le 1$.

Covariance between periodogram ordinates

Below we establish the main result of this section which calculates the covariance between periodogram ordinates of realizations X_0^{n-1} and Y_0^{n-1}:

Lemma 2.14 *If X_0^{n-1} and Y_0^{n-1} are given by the MC with transfer functions F and G, then*

$$\begin{aligned}\operatorname{cov}(I_n^x(\lambda), I_n^y(\mu)) &= n^{-2}\left|\int F(\omega)\overline{G(\omega)}D_n(\omega-\lambda)D_n(\omega-\mu)d\omega\right|^2 \\ &+ n^{-2}\left|\int F(\omega)\overline{G(\omega)}D_n(\omega-\lambda)D_n(\omega+\mu)d\omega\right|^2.\end{aligned}$$

In the case $F = G$, i.e., $X_t = Y_t$, lemma 2.14 leads to the following known result (cf., e.g., Haszminski and Ibragimov, 1986):

Corollary 2.15 *If $X_0^{n-1} \sim N(0, f)$, then*

$$\begin{aligned}\operatorname{cov}(I_n^x(\lambda), I_n^x(\mu)) &= n^{-2}\left|\int f(\omega)D_n(\omega-\lambda)D_n(\omega-\mu)d\omega\right|^2 \\ &+ n^{-2}\left|\int f(\omega)D_n(\omega-\lambda)D_n(\omega+\mu)d\omega\right|^2 \qquad (2.9)\end{aligned}$$

It is well known that the components of $I_n^x(k/n)$, $k = 0, \ldots, [n/2]$ become increasingly uncorrelated as $n \to \infty$. Since spectral estimates usually depend only on these components, the properties of these estimates may depend on the rate at which the correlation between these components decays to 0. For this reason, it is useful to evaluate (2.9) explicitly when λ and μ are different Fourier frequencies. We then have the following

Lemma 2.16 *If $X_0^{n-1} \sim N(0, f)$, $f \in \operatorname{Lip}(1)$, and $0 \le l < m \le [n/2]$, then*

$$\operatorname{cov}(I_n^x(l/n), I_n^x(m/n)) = M_f^2 O(\log n/n)^2.$$

uniformly in f.

We conclude this section with a result on the magnitude of the covariance between the logs of periodogram ordinates. This result will prove useful when establishing properties of the particular spectral estimate (Wahba, 1980) used in our simulations since this procedure essentially amounts to taking a weighted average of the log of periodogram ordinates.

Lemma 2.17 *If X_0^{n-1} satisfies the conditions of Definition 2.3 and $0 \le l < m \le [n/2]$, then*

$$\operatorname{cov}(\log I_n^x(l/n), \log I_n^x(m/n)) = M_f^2 O(\log n/n)^2. \qquad (2.10)$$

uniformly in f.

2.4 Circulants

In this section we introduce circulants, describe some of their properties, and discuss in some detail two examples of circulants which will appear later on.

Definition 2.8 *A matrix $\Sigma_n^c = (\sigma_{kj})$, $k, j = 0, \ldots, n-1$, is a circulant of dimension n if $\sigma_{kj} = \sigma^c(k-j)$ for some function σ^c with period n.*

Recall (cf., e.g., Brillinger, 1979) that the eigenvalues of Σ_n^c are given by

$$\lambda_j^c = \sum_{k=0}^{n-1} \sigma^c(k) \exp(-i2\pi kj/n),$$

$j = 0, ..., n-1$, and the corresponding right-hand eigenvectors are simply the rows (or columns) of the Fourier transform matrix F_n. In fact, it is well known (cf. e.g. Brillinger, 1979) that Σ_n^c is a circulant if, and only if, it can be written as

$$\Sigma_n^c = F_n \Lambda_n^c \overline{F_n} \tag{2.11}$$

where Λ_n^c is diagonal with entries λ_k^c. The importance of (2.11) is that the eigenvectors, the columns of F_n, are the same for all circulants of dimension n. Notice that whenever Σ_n^c and Λ_n^c are both real, Σ_n^c is necessarily symmetric (and hence Toeplitz) and $\Sigma_n^c = \overline{F_n} \Lambda_n^c F_n$.

It is easy to see that, given the eigenvalues, we can obtain the entries of Σ_n^c by taking the inverse DFT

$$\sigma^c(k) = \frac{1}{n} \sum_{j=0}^{n-1} \lambda_j^c \exp(i2\pi jk/n). \tag{2.12}$$

2.4.1 Example 1: A circulant approximation to Σ_n

Let X_0^{n-1} be a realization of size n from $N(0, f)$, $f > 0$, and let Σ_n denote its covariance matrix, i.e., $X_0^{n-1} \sim N(0, \Sigma_n)$, $\Sigma_n = T(c_0, \ldots, c_{n-1})$. If we let $J_n^x = F_n X$ denote its DFT, it follows from Lemma 2.11 that

$$E J_n^x \overline{J_n^x}^T = F_n \Sigma_n \overline{F_n} = \Lambda_n.$$

so that

$$\Sigma_n = \overline{F_n} \Lambda_n F_n = F_n \overline{\Lambda_n} \overline{F_n}.$$

where Λ_n is not necessarily diagonal or real. However,

Lemma 2.18 *The diagonal entries of Λ_n are given by*

$$\lambda_j = \sum_{t=-n+1}^{n-1} c_t (1 - |t|/n) \exp(-i2\pi tj/n) \tag{2.13}$$

$$= \sum_{t=0}^{n-1} (c_t (1 - |t|/n) + c_{n-t}|t|/n) \exp(-i2\pi tj/n) \tag{2.14}$$

Remark: The $\lambda_j = K_n * f(j/n)$ are the Cesaro means of f evaluated at the Fourier frequencies j/n, $j = 0, \ldots, n-1$.

In order to construct a circulant approximation to Σ_n, define $\Lambda_n^c = \text{diag}(\Lambda_n)$ to be the diagonal matrix whose entries coincide with the entries in the main diagonal of Λ_n and let

$$\Sigma_n^c = F_n \Lambda_n^c \overline{F_n} = \overline{F_n} \Lambda_n^c F_n.$$

Lemma 2.12 shows that the off-diagonal entries of Λ_n are $O(\log n/n)$ so that Σ_n^c is a circulant approximation to Σ_n.

By construction it is clear that Σ_n^c is a circulant since it is in the form of (2.11). The entries of Λ_n^c are given by (2.14) and hence the inversion formula (2.12) gives

$$c_k^c = c_k (1 - k/n) + c_{n-k} k/n. \tag{2.15}$$

for $k = 0, \ldots, n-1$. Notice that the approximation is better for entries closer to the main diagonal.

2.4.2 Example 2: "Circularizing" processes

Let $f(\omega) = \sum_s c_s \exp(i2\pi\omega s)$ be a spectrum and consider the following $2n \times 2n$ circulant

$$\Sigma^c_{2n} = T(c_0, \ldots, c_{n-1}, 0, c_{n-1}, \ldots, c_1).$$

Lemma 2.19 *If $\Sigma^c_{2n} = F_{2n}\Lambda^c_{2n}\overline{F}_{2n}$, then Λ^c_{2n} is diagonal with entries*

$$\lambda^c_j = \sum_{t=-n+1}^{n-1} c_t \exp(i2\pi tj/2n)$$

for $j = 0, \ldots, 2n-1$.

Remarks:

1. The eigenvalues $\lambda^c_j = D_{2n+1} * f(j/2n)$ are the partial sum of the Fourier series of f evaluated at the Fourier frequencies $j/2n$, for $j = 0, \ldots, 2n-1$. Since the Dirichlet kernel is not positive everywhere it may be that some λ^c_j's are negative even when $f > 0$, and hence that Σ^c_{2n} is not necessarily positive definite.

2. If $\lambda_j > 0$ for all j, then Σ^c_{2n} is positive definite and hence is the covariance matrix of a circular process with period $2n$. This process has the property that any n adjacent components of a realization are distributed $N(0, \Sigma_n)$ where $\Sigma_n = T(c_0, \ldots, c_{n-1})$.

The last property will be useful for fast simulation of realizations from $N(0, \Sigma_n)$ since the FFT can be used to generate realizations of processes which have circulant covariance matrices. For this reason, it will be important that for large enough n, Σ^c_{2n} be positive definite, i.e., that it be a covariance matrix. This is guaranteed by the following

Lemma 2.20 *If $\sum |c_k| < \infty$ and $f > 0$, then there exists N such that Σ^c_{2n} is positive definite for $n > N$.*

Chapter 3

Linear functionals of the spectrum

3.1 Introduction

Recall the motivating problem stated in the introduction: upon observing a realisation $X_0^{n-1} \sim N(0, f)$, f unknown, procedure $\hat{\theta}_x$ is used to estimate a functional θ_f of f and it is of interest to assess the sampling distribution of $\hat{\theta}_x$.

One possible solution to this problem is the following: use some procedure $\hat{f}_n$ to estimate f and, by simulating many realizations from a process $\{Y_t\}$ whose conditional distribution given $\hat{f}_n$ is $N(0, \hat{f}_n)$, determine the (conditional) sampling distribution of $\hat{\theta}_y$ and use it to estimate that of $\hat{\theta}_x$. An obvious question is: under what conditions on θ, $\hat{\theta}$, f, and $\hat{f}_n$, can the sampling distribution of $\hat{\theta}_x$ be consistently estimated by the (conditional) sampling distribution of $\hat{\theta}_y$?

In this chapter we investigate the answer to this question in the case where the parameter of interest is a linear functional of the spectrum and its estimate is the corresponding functional of the periodogram. The main result in this chapter, Corollary 3.5, establishes that in this case, under certain conditions on $\hat{f}_n$ and θ_f, $\mathcal{L}(\hat{\theta}_x)$ and $\mathcal{L}(\hat{\theta}_y|X_0^{n-1})$ are close with high probability. This result is then extended to well-behaved functions of linear functionals in Corollary 3.9. For example, we shall be interested in assessing the sampling distribution of estimators of normalised linear functionals such as the ratio of two linear functionals.

Our results will, of course, depend on properties of the particular estimator of the spectral density used. In Chapter 4 we show that the estimator of the spectral density (Wahba, 1980) which we use in our simulations, will have the required properties whenever the smoothing parameter is properly chosen.

Some of the theory required for establishing these results can be found in §A.2.1.

3.2 Sampling distribution of estimators of linear functionals

3.2.1 Efficient estimators

Assume we observe a realisation $X_0^{n-1} \sim N(0, f)$, f unknown, and we estimate the parameter

$$\theta_f = \int \theta(\omega) f(\omega) d\omega \tag{3.1}$$

by

$$\hat{\theta}_x = \int \theta(\omega) I_n^x(\omega) d\omega \tag{3.2}$$

for a real and measurable function θ. There is no loss of generality in assuming that θ is even and periodic and we shall hereafter do so.

Linear functionals include a wide class of par- eters, e.g., covariances and spectral moments. In addition, they have nice properties: Hasminskii and Ibragimov (1986) establish (under weaker conditions on f and θ than those we assume in this paper) that $\hat{\theta}_x$ is asymptotically efficient, and

$$\mathcal{L}\left(\sqrt{n}(\hat{\theta}_x - \theta_f)\right) \to \mathcal{L}(U) \tag{3.3}$$

where $U \sim N\left(0, 2\int \theta(\omega)^2 f(\omega)^2 d\omega\right)$.

3.2.2 Resampling strategy

We saw above that $\sqrt{n}(\hat{\theta}_x - \theta_f)$ has a limiting distribution. For this reason it will be easier to deal with the sampling distribution of

$$U_n = \sqrt{n}(\hat{\theta}_x - \theta_f), \tag{3.4}$$

rather than that of $\hat{\theta}_x$.

The idea behind our resampling methodology is to estimate f by $\hat{f}_n$ and estimate the distribution of U_n by the *conditional distribution given* $\hat{f}_n = g$ of

$$V_n = \sqrt{n}(\hat{\theta}_y - \theta_g), \tag{3.5}$$

where $\mathcal{L}(Y_0^{n-1}|X_0^{n-1}) = N(0, \hat{f}_n)$.

3.2.3 Results

The results of this section are obtained by showing that there exist random variables $\tilde{U}_n$ and $\tilde{V}_n$ which are close with high probability and such that $\mathcal{L}(\tilde{U}_n|X_0^{n-1}) = \mathcal{L}(\tilde{U}_n) = \mathcal{L}(U_n)$ and $\mathcal{L}(\tilde{V}_n|X_0^{n-1}) = \mathcal{L}(V_n|X_0^{n-1})$, i.e., $\tilde{U}_n - \tilde{V}_n = o_p(1)$. Corollary 3.5 then establishes

1. $d_P\left(\mathcal{L}(U_n|X_0^{n-1}), \mathcal{L}(\tilde{V}_n|X_0^{n-1})\right) = o_p(1)$ and
2. $d_P\left(\mathcal{L}(U), \mathcal{L}(\tilde{V}_n|X_0^{n-1})\right) = o_p(1)$.

where d_P denotes the Prohorov distance (cf., e.g., §A.2.1). In other words, the distribution of U_n is approximated by the "bootstrap" distribution.

In the next section we construct these random variables and show that they indeed have the desired properties. To do so, we proceed in stages: first, we assume $\hat{f}_n = g$ is given and deterministic and use the MC to construct processes $\{\tilde{X}_t\} \sim N(0, f)$ and $\{\tilde{Y}_t\} \sim N(0, g)$ and define $\tilde{U}_n$ and $\tilde{V}_n$ as in (3.4) and (3.5); second, we bound the distance between the distributions of $\tilde{U}_n$ and $\tilde{V}_n$ in terms of properties of f and g; third, we extend the results to account for the fact that $\hat{f}_n$ is really a random variable by determining regularity conditions on $\hat{f}_n$ so that the required properties of deterministic g hold for random $\hat{f}_n$ with high probability.

3.3 Distances between distributions of estimators

3.3.1 Deterministic case

If $\hat{f}_n = g$ is deterministic, the basic idea is to study how "close" the distributions of U_n and V_n are as a function of how "close" the respective spectra f and g are. It turns out (Theorem 3.3) that if the two processes $\{X_t\} \sim N(0, f)$ and $\{Y_t\} \sim N(0, g)$ are given by the MC and we measure the distance between the distributions of U_n and V_n by $E|U_n - V_n|$, this distance is essentially bounded by the L_2 distance between f and g.

Before we go on we need to establish two auxiliary results. The first one concerns the bias of $\hat{\theta}_x$.

Lemma 3.1 *If θ is bounded, $\hat{\theta}_x$ is given by (3.2), $f \in \mathrm{Lip}(1)$, and $X_0^{n-1} \sim N(0, f)$, then*

$$E(\hat{\theta}_x - \theta_f) = M_f\, O(\log n/n)$$

uniformly in f.

The second one concerns the covariance between $\hat{\theta}_x$ and $\hat{\phi}_y = \int \phi(\omega) I_n^y(\omega) d\omega$, an estimate of $\phi_g = \int \phi(\omega) g(\omega) d\omega$.

Lemma 3.2 *If $X_0^{n-1} \sim N(0, f)$ and $Y_0^{n-1} \sim N(0, g)$ are given by the MC with transfer functions F and G, and $\theta, \phi \in \mathrm{Lip}(1)$ are even and periodic, then*

$$n\mathrm{cov}(\hat{\theta}_x, \hat{\phi}_y) = 2 \int \theta(\omega)\phi(\omega) f(\omega) g(\omega) d\omega + M_{F\overline{G}} \sup |FG| O(\log n/n) + \sup |fg| O(\log^2 n/n)$$

uniformly in F and G.

We are now able to establish the establish the main result of this section, Theorem 3.3, which bounds the distance between the distribution of U_n and V_n by $E|U_n - V_n|$ where U_n and V_n are induced by processes $\{X_t\}$ and $\{Y_t\}$ given by the MC.

Theorem 3.3 *If $X_0^{n-1} \sim N(0, f)$ and $Y_0^{n-1} \sim N(0, g)$ are given by the MC with transfer functions F and G, $\theta \in \mathrm{Lip}(1)$ is even and periodic, and U_n and V_n are given by (3.4) and (3.5), then*

$$\begin{aligned} E^2|U_n - V_n| &\leq E|U_n - V_n|^2 \\ &= 2 \int \theta(\omega)^2 |f(\omega) - g(\omega)|^2 d\omega \\ &\quad + M_{F\overline{G}} \sup |FG| O(\log n/n) + (\sup |fg| + M_f^2 + M_g^2) O(\log^2 n/n) \end{aligned} \tag{3.6}$$

uniformly in F and G.

This result will enable us to show that $U_n - V_n = o_p(1)$ when g is no longer deterministic but a consistent (in integrated mean square error) estimate $\hat{f}_n$ with certain regularity properties. Consistency ensures the integral in (3.6) to be $o_p(1)$ while the regularity conditions ensure the remaining terms in (3.6) to be $o_p(1)$.

3.3.2 Random case

Extending the results of the previous section to the random case where $\hat{f}_n$ is random will follow upon showing that the estimate of the spectral density based on the original sample X_0^{n-1} and the corresponding estimate of the transfer function satisfy certain properties which we will now explore.

Let $\hat{f}_n$ denote the estimate of the spectral density which the observed data X_0^{n-1} induce. We will use the MC with $F = \sqrt{f}$, $G = \{\hat{f}_n\}^{1/2}$ and $Z(\omega)$ independent of X_0^{n-1} to generate realizations of two processes $\{\tilde{X}_t\}$ and $\{\tilde{Y}_t\}$ whose conditional distributions given X_0^{n-1} (or equivalently $\hat{f}_n$) are $N(0, f)$ and $N(0, \hat{f}_n)$. These processes in turn induce random variables $\tilde{U}_n$ and $\tilde{V}_n$ with the property that $\mathcal{L}(\tilde{U}_n|X_0^{n-1}) = \mathcal{L}(\tilde{U}_n) = \mathcal{L}(U_n)$ and $\mathcal{L}(\tilde{V}_n|X_0^{n-1}) = \mathcal{L}(V_n|X_0^{n-1})$. We will show that under regularity conditions on $\hat{f}_n$ these two distributions are close with high probability.

Theorem 3.4 *Let $\hat{f}_n$ be a spectral estimate which depends on X_0^{n-1} and satisfies the following properties:*

1. $\|f - \hat{f}_n\|_2 = o_p(1)$
2. $P(\inf_\omega \hat{f}_n(\omega) \geq 0) \to 1$.
3. $\sup_\omega \hat{f}_n(\omega) = O_p(1)$
4. $\sup_\omega \hat{f}_n^{-1}(\omega) = O_p(1)$
5. $\sup_\omega \hat{f}_n'(\omega)^2 = o_p(n/\log^2 n)$.

If $\theta \in \mathrm{Lip}(1)$ is even and periodic, then

$$\tilde{U}_n - \tilde{V}_n = o_p(1).$$

Corollary 3.5 *Let $U \sim N(0, 2\int \theta(\omega)^2 f(\omega)^2 d\omega)$. If the conditions of Theorem 3.4 hold, then*

1. $d_P\left(\mathcal{L}(U_n), \mathcal{L}(\tilde{V}_n|X_0^{n-1})\right) = o_p(1)$ *and*
2. $d_P\left(\mathcal{L}(U), \mathcal{L}(\tilde{V}_n|X_0^{n-1})\right) = o_p(1)$.

Remark: Item (1) above shows that the sampling distribution of U_n is consistently estimated by the "bootstrap" distribution.

3.4 Functions of linear functionals

In this section we extend the results of the previous section to well-behaved functions of linear functionals of the spectrum, i.e., the case where

$$h(\hat{\theta}_{x1}, \ldots, \hat{\theta}_{xp}) = h(\hat{\theta}_x)$$

is used to estimate

$$h(\theta_{f1}, \ldots, \theta_{fp}) = h(\theta_f)$$

for some function h which is well-behaved.

3.4.1 Limiting distribution

In this section we establish that the limiting distribution of well-behaved functions of linear functionals is multivariate Gaussian. First, we show that the limiting distribution of vectors of linear functionals is multivariate Gaussian.

Lemma 3.6 *Let $f, \theta_j \in \text{Lip}(1)$, $j = 1, \ldots, p$, and let*

$$\hat{\theta}_x' = (\hat{\theta}_{x1}, \ldots, \hat{\theta}_{xp})'$$

be a vector of estimates of

$$\theta_f' = (\theta_{f1}, \ldots, \theta_{fp})'$$

where

$$\hat{\theta}_{xj} = \int \theta_j(\omega) I_n^x(\omega) d(\omega)$$

and

$$\theta_{fj} = \int \theta_j(\omega) f(\omega) d(\omega)$$

for $j = 1, \ldots, p$. Then, the limiting distribution of $\sqrt{n}(\hat{\theta}_x - \theta_f)$ is $N(0, \Sigma)$ where

$$\Sigma = 2 \left\| \int \theta_j(\omega) \theta_k(\omega) f(\omega) d\omega \right\|.$$

The following theorem can then be used to establish asymptotic normality of estimators well-behaved functions of linear functionals.

Theorem 3.7 *(Rao, 1972) If $T_n = (T_{1n}, \ldots, T_{kn})$ is a k-dimensional random variable such that the asymptotic distribution of*

$$\begin{pmatrix} \sqrt{n}(T_{1n} - \theta_1) \\ \vdots \\ \sqrt{n}(T_{kn} - \theta_k) \end{pmatrix}$$

is k-variate normal with mean zero and covariance matrix $\Sigma = (\sigma_{ij})$ and h is a function of k variables which is totally differentiable, then the asymptotic distribution of

$$\sqrt{n}(h(T_{1n}, \ldots, T_{kn}) - h(\theta_1 \ldots \theta_k))$$

is normal with mean zero and variance

$$v(\theta) = \sum_i \sum_j \sigma_{ij} \frac{\partial h}{\partial \theta_i} \frac{\partial h}{\partial \theta_j}$$

provided $v(\theta) > 0$, $\theta = (\theta_1, \ldots, \theta_k)'$.

3.4.2 Results

In this section we extend the result of Theorem 3.4 to well-behaved functions of linear functionals of the spectrum.

Theorem 3.8 *Let $\hat{f}_n$ be an estimate of the spectrum f which depends on X_0^{n-1} and satisfies the assumptions of Theorem 3.4. Let $\tilde{X}_0^{n-1}$ and $\tilde{Y}_0^{n-1}$ be realizations of two processes given by the MC with transfer functions $F = (f)^{1/2}$ and $G = (\hat{f}_n)^{1/2}$, and let $\theta_j \in \text{Lip}(1)$, $j = 1, \ldots, p$ be even and periodic. If*

$$\tilde{U}_n^* = \sqrt{n}(h(\hat{\theta}_{\tilde{x}}) - h(\theta_f)), \; \tilde{V}_n^* = \sqrt{n}(h(\hat{\theta}_{\tilde{y}}) - h(\theta_{f_n})),$$

and h has continous derivatives in some neighborhood N_{θ_f} of θ_f, then

$$\tilde{U}_n^* - \tilde{V}_n^* = o_p(1).$$

Corollary 3.9 *Assume the conditions of Theorem 3.8 hold, and let U^* be a random variable whose distribution is the same as the limiting distribution of $\tilde{U}_n^*$. Then*

1. $a_P\left(\mathcal{L}(U_n^*), \mathcal{L}(\tilde{V}_n^* | X_0^{n-1})\right) = o_p(1)$ *and*
2. $d_P\left(\mathcal{L}(U^*), \mathcal{L}(\tilde{V}_n^* | X_0^{n-1})\right) = o_p(1)$.

3.4.3 Application: normalized functionals

We illustrate the results of the previous section with an example which we use later on in our simulations.

Theorem 3.7 can be used to establish asymptotic normality of estimates of normalized linear functionals, i.e., estimates

$$\hat{\theta}_x^* = \frac{\int \theta(\omega) I_n^x(\omega) d\omega}{\int I_n^x(\omega) d\omega} = \frac{\hat{\theta}_x}{\|I_n^x\|_1}. \tag{3.7}$$

of functionals

$$\theta_f^* = \frac{\int \theta(\omega) f(\omega) d\omega}{\int f(\omega) d\omega} = \frac{\theta_f}{\|f\|_1}, \tag{3.8}$$

The class of normalized functionals includes the correlations

$$\rho_k = \frac{\int cos(2\pi\omega) f(\omega) d\omega}{\int f(\omega) d\omega}$$

and the normalized spectral moments.

A corollary to Theorem 3.7 gives the form of the limiting distribution of normalized linear functionals.

Corollary 3.10 *If $U_n = \sqrt{n}(\hat{\theta}_x^* - \theta_f^*)$, then its limiting distribution is*

$$N\left(0, 2(\theta_f^*)^2 \int (\theta(\omega)/\theta_f^* - 1)^2 f^*(\omega)^2 d\omega\right).$$

where $f^* = f/\|f\|_1$.

Corollary 3.9 implies that the sampling distribution of U_n can be consistently estimated by our resampling methodology. In Chapter 6 we give the results from simulations where the parameter of interest is a normalized functional, the first correlation.

Chapter 4

Estimating the spectral density

4.1 Wahba's estimate

In this chapter we introduce Wahba's (1980) estimate and establish that it has the properties which will allow us to apply the results of Chapter 3 when it is used as the spectral density to generate the bootstrap samples.

The procedure estimates the log-spectrum of a stationary process with positive and differentiable spectrum $f \in \text{Lip}(1)$. In this case the log-spectrum $h \in \text{Lip}(1)$ and has a Fourier series expansion, i.e.,

$$h(\omega) = \log f(\omega) = \sum_{k=-\infty}^{\infty} b_k \exp(i2\pi\omega k).$$

We will assume, in addition, that $\sum_{k=-\infty}^{\infty} |b_k| < \infty$.

Now, suppose we observe a realization of length $2n$ from a Gaussian stationary process with spectrum f, i.e., we observe $X_0^{2n-1} \sim N(0, f)$. The proposed estimate of $h(\omega)$ is

$$\hat{h}_n(\omega) = \sum_{\nu=-n}^{n} \frac{\hat{b}_{\nu n}}{1 + \lambda(2\pi\nu)^{2m}} \exp(i2\pi\nu\omega), \tag{4.1}$$

where

$$\hat{b}_{\nu n} = \frac{1}{2n} \sum_{k=-n+1}^{n} Y_{kn} \exp(-i2\pi\nu k/2n), \tag{4.2}$$

$Y_{kn} = \log I_{2n}^{n}(k/2n) + \gamma$, and $\gamma = .57721\ldots$ is Euler's constant.

From (4.1) and (4.2) it is easy to show that $\hat{h}_n$ is a smoothed version of the log-periodogram since it can be rewritten as

$$\hat{h}_n(\omega) = \sum_{k=-n+1}^{n} Y_{kn} W_{\lambda,m,n}(\omega - k/2n) \tag{4.3}$$

where

$$W_{\lambda,m,n}(\omega) = \frac{1}{2n} \sum_{\nu=-n}^{n} \frac{\exp(i2\pi\nu\omega)}{1 + \lambda(2\pi\nu)^{2m}} \tag{4.4}$$

is real, periodic, not necessarily positive, and such that

$$\int_0^1 W_{\lambda,m,n}(\omega)d\omega = \sum_{k=-n+1}^{n} W_{\lambda,m,n}(\omega - k/2n) = 1.$$

4.2 Properties of the log-periodogram

In order to study the properties of (4.3) it is convenient to first assess the properties of the triangular array Y_{kn}, $k = -n+1, \ldots, n$. We can write

$$\begin{aligned} Y_{kn} &= \log\left(K_{2n} * f(k/2n)V_{kn}\right) + \gamma \\ &= \log(K_{2n} * f(k/2n)) + \epsilon_{kn} \end{aligned}$$

where $\epsilon_{kn} = \log V_{kn} + \gamma$, and establish the following

Lemma 4.1 *If $f \in Lip(1)$ and I^x_{2n} is the periodogram from a realization of size $2n$ from $N(0, f)$, then*

1. $EY_{kn} = \log(K_{2n} * f(k/2n)) + M_f^2 O(\log n/n)^2$
2. $E\epsilon_{kn} = M_f^2 O(\log n/n)^2$
3. $\mathrm{Var}Y_{kn} = \mathrm{Var}(\epsilon_{kn}) = \pi^2/6 + M_f^2 O(\log n/n)^2$
4. $\mathrm{cov}(Y_{kn}, Y_{jn}) = \mathrm{cov}(\epsilon_{kn}, \epsilon_{jn}) = M_f^2 O(\log n/n)^2$

uniformly in f for $1 \le k < j \le n-1$.

4.3 Properties of Wahba's estimate

In this section we derive properties of Wahba's estimate which allow us to apply the results of Chapter 3 when we use this estimate of the spectrum, in other words, we show that Wahba's estimate has the regularity properties which Theorem 3.4 requires. We introduce notation and auxiliary results first.

The following lemma derives properties of the window $W_{\lambda,m,n}$ and its derivatives.

Lemma 4.2 *If $W_{\lambda,m,n}(\omega)$ is given by (4.4), $0 \le k < 2m$, and*

$$W^{(k)}_{\lambda,m,n}(\omega) = \frac{d^k}{d\omega^k} W_{\lambda,m,n}(\omega),$$

then

$$\sup_\omega \left| W^{(k)}_{\lambda,m,n}(\omega) \right| = (\lambda^{-k/2m} + \lambda^{-(1+k)/2m}) O(n^{-1})$$

uniformly in $\lambda > 0$.

Remark: There are two terms involving powers of λ in the expression above to cover $\lambda \to 0$ and $\lambda \to \infty$. In this paper we will usually need to use this result when $\lambda \to 0$ in which case Lemma 4.2 implies

$$\sup_\omega \left| W^{(k)}_{\lambda,m,n}(\omega) \right| = \lambda^{-(1+k)/2m} O(n^{-1}).$$

Now, notice that, for n large, the estimate (4.3) of $h = \log f$ can be rewritten as

$$\begin{aligned} \hat{h}_n(\omega) &= \sum_{k=-n+1}^{n} \left(\log(K_{2n} * f(k/2n) + \epsilon_{kn}\right) W_{\lambda,m,n}(\omega - k/2n) \\ &= \sum_{k=-n+1}^{n} W_{\lambda,m,n}(\omega - k/2n) \log(K_{2n} * f(k/2n)) \\ &\quad + \sum_{k=-n+1}^{n} W_{\lambda,m,n}(\omega - k/2n)\epsilon_{kn} \\ &= \tilde{h}_n(\omega) + h^*_n(\omega) \end{aligned} \tag{4.5}$$

so that it consists of a deterministic term and a random term.

4.3.1 Properties of the deterministic term

In this section we find bounds for the deterministic term and its first derivative. In addition, we show that it converges in L_2 norm to the true log-spectrum.

Lemma 4.3 $\sup_\omega |\tilde{h}_n(\omega)| \le \sum_{\nu=-\infty}^{\infty} |b_\nu| + (1+\lambda^{-1/2m})O(\log n/n)$ *uniformly in* λ.

Lemma 4.4 *If* $f \in \mathrm{Lip}(1)$ *is differentiable, then*

$$\sup_\omega |\tilde{h}_n'(\omega)| = (1+\lambda^{-1/2m})O(1)$$

uniformly in λ.

The following lemma establishes convergence to h in L_2.

Lemma 4.5 *If* $f \in \mathrm{Lip}(1)$ *and* $\tilde{h}_n$ *is given as in (4.5), then*

$$\|h - \tilde{h}_n\|_2 \le A(\lambda) + (1+\lambda^{-1/2m})\,O(\log n/n) \tag{4.6}$$

uniformly in λ, *where*

$$A^2(\lambda) = \sum_{\nu=-\infty}^{\infty} \left(\frac{b_\nu \lambda (2\pi\nu)^{2m}}{1+\lambda(2\pi\nu)^{2m}} \right)^2 \tag{4.7}$$

so that $A(\lambda)$ *is bounded, continuous, and* $A(0) = 0$.

4.3.2 Properties of the random term

In this section we establish bounds for the random term and its first derivative, and its integrated mean square. First, the result for the integrated mean square.

Lemma 4.6 *If* $f \in \mathrm{Lip}(1)$ *and* h_n^* *is given as (4.5), then*

$$E \int h_n^*(\omega)^2 d\omega = O(n^{-1})(1+\lambda^{-1/2m}).$$

uniformly in λ.

The following result establishes the regularity properties for the random term.

Lemma 4.7 *Let* $h_n^*(\omega) = \sum_{k=-n+1}^{n} \epsilon_{kn} W_n(\omega)$ *and let the following conditions hold:*

1. $E\epsilon_{kn} = O(\log^2 n/n^2)$
2. $\sup_k E|\epsilon_{kn}|^2 = O(1)$
3. $\sup_{k,j} E\epsilon_{kn}\epsilon_{jn} = O(\log^2 n/n^2)$
4. $\sum_k W_n(\omega - k/2n) = 1$
5. $\sup_\omega |W_n(\omega)| = O(n^{r_0-1})$
6. $\sup_\omega |W_n'(\omega)| = O(n^{r_1-1})$
7. $\sup_\omega |W_n''(\omega)| = O(n^{r_2-1})$,

where $0 \le r_i < 1$, $i = 0,1,2$. *Then, for any* $0 < \delta, s < 1$, *we have that*

$$\sup_\omega |h_n^*(\omega)| \le O_p(n^{1/2+\delta+r_0-s} + n^{s+r_1-1}) \tag{4.8}$$

and

$$\sup_\omega \left| \frac{d}{d\omega} h_n^*(\omega) \right| \le O_p(n^{1/2+\delta+r_1-s} + n^{s+r_2-1}). \tag{4.9}$$

This result requires interpretation: Assume that $W_n = W_{\lambda,m,n}$ where the smoothing sequence λ_n satisfies $\lambda_n = n^{-r}$. Then $r_0 = r/2m$, $r_1 = r/m$, and $r_2 = 3r/2m$. Now let

$$\alpha_1 = \max\left(\frac{1}{2} + \delta + \frac{r}{2m} - s, s + \frac{r}{m} - 1\right)$$

and

$$\alpha_2 = \max\left(\frac{1}{2} + \delta + \frac{r}{m} - s, s + \frac{3r}{2m} - 1\right).$$

Then, if we take $s = \frac{3}{4} + \frac{\delta}{2} - \frac{r}{4m}$ and $0 < r < m/3$, we have that $\alpha_1 = 1/4 + 3r/4m + \delta/2 < \delta/2$ and $\alpha_2 = -1/4 + 5r/4m + \delta/2 < \delta/2 + 1/6$ so that, for δ sufficiently small, it follows that $\alpha_1 \leq 0$, $\alpha_2 \leq 1/6$, and hence, $\sup|h_n^*| = O_p(1)$, and $\sup|h_n^{*\prime}| = O_p(n^{1/6})$.

4.3.3 General properties

We can now establish a result for the integrated mean square error of $\hat{h}_n$.

Lemma 4.8 *If $f \in \mathrm{Lip}(1)$ and $\hat{h}_n$ is given by (4.5), then*

$$Ed_2(h, \hat{h}_n) = A(\lambda) + (1 + \lambda^{-1/2m})\, O(\log n/n) + O(n^{-1/2})(1 + \lambda^{-1/4m}). \tag{4.10}$$

uniformly in λ.

If $\hat{f}_n = \exp(\hat{h}_n)$ is used to estimate the spectrum f, we have

Lemma 4.9 *Let $f \in \mathrm{Lip}(1)$ and $\lambda_n = Kn^{-r}$ for $0 < r < m/3$. Then*

$$d_2(\hat{f}_n, f) = O_p(1)\left(O(\log n/n^{1-r/2m}) + A(Kn^{-r}) + O(n^{-1/2})(1 + \lambda^{-r/4m})\right).$$

uniformly in λ.

4.4 Summary

We have established that, if $f \in \mathrm{Lip}(1)$ is differentiable, $\sum|b_\nu| < \infty$, and $\lambda_n = O(n^{-r})$, $0 < r < m/3$, the estimate of the spectral density given by $\hat{f}_n = \exp(\hat{h}_n)$ has the following properties:

1. $\sup_\omega |h_n(\omega)| = O_p(1)$ (Lemmas 4.3 and 4.7) which implies
   (a) there exists $\epsilon > 0$ such that $P(\inf_\omega \hat{f}_n(\omega) > \epsilon) \to 1$ and
   (b) $\sup_\omega f_n(\omega) = O_p(1)$
2. $\sup_\omega |h'(\omega)| = O_p(n^{1/6})$ (Lemmas 4.4 and 4.7) which implies $\sup_\omega |f'(\omega)| = O_p(n^{1/6})$
3. $\|\hat{f}_n - f\|_2 = o(1)$ (Lemma 4.9).

These are sufficient for the conditions of Theorem 3.4 to obtain.

Chapter 5

Simulation algorithms

5.1 Introduction

Recall the motivating problem stated in the introduction: a realisation $X_0^{n-1} \sim N(0, f)$, f unknown, is observed, procedure $\hat{\theta}_x$ is used to estimate a functional θ_f of f, and it is of interest to estimate the sampling distribution of $\hat{\theta}_x$.

In chapter 3 we proposed the following approach to this problem: use some procedure $\hat{f}_n$ to estimate f and, by simulating many realisations from a process $\{Y_t\}$ whose conditional distribution given $\hat{f}_n$ is $N(0, \hat{f}_n)$, determine the (conditional) sampling distribution of $\hat{\theta}_y$ and use it to estimate that of $\hat{\theta}_x$.

From a practical point of view the problem is two-fold: (i) finding an estimate $\hat{f}_n$ of f, and (ii) calculating the sampling distribution of $\hat{\theta}_y$, $Y \sim N(0, \hat{f}_n)$. The first problem was addressed in the previous chapter where we introduced Wahba's (1980) estimator of the log-spectrum and discussed its properties. As far as the second problem is concerned, our approach will be to use Monte-Carlo to approximate the distribution of $\hat{\theta}_y$ by the empirical cdf of $\hat{\theta}_y$ for many realisations $Y_0^{n-1} \sim N(0, \hat{f})$.

The Monte-Carlo approach to estimating distributions makes generating simulations quickly very desirable. For this reason, one of the emphases of this research was to implement fast algorithms to simulate realisations $Y_0^{n-1} \sim N(0, g)$ for a given g.

5.2 Simulating Gaussian stationary processes

There are several ways to simulate realisations from $N(0, g)$. For example, if we let $c_0, \ldots, c_{n-1}$ denote the first n terms in the Fourier series expansion of g, then (Lemma 2.1) $\Sigma_n = T(c_0, \ldots, c_{n-1})$ is positive-definite whenever $g > 0$. It is thus simple, at least in principle, to obtain a realization from $N(0, \Sigma_n)$: let $Y_0^{n-1} = \Sigma_n^{1/2} Z_0^{n-1}$ where Z_0^{n-1} is a vector of iid $N(0, 1)$ variates.

The problem with this approach lies in calculating $\Sigma_n^{1/2}$. Obtaining the square root of an arbitrary positive definite (p.d.) matrix requires $O(n^3)$ operations. Depending on the amount of structure in the matrix this calculation can be speeded up. For example, by exploiting the Toeplits nature of Σ_n, Carlin et al. (1983) describe an algorithm (essentially the Levinson recursion) to obtain a Cholesky factorisation of Σ_n in $O(n^2)$ operations.

The question is whether we can devise an algorithm which requires less than $O(n^2)$ operations to generate realisations from $N(0, g)$ by operating in the frequency domain and taking advantage of the FFT's speed. It so happens that there is an algorithm which only requires $O(n \log n)$ operations. The constant in O will depend on the smoothness of g.

5.2.1 Simulation of circular processes

In order to explain our approach it is easier to start with the simplest case. It turns out that this is the case where the covariance matrix is a circulant Σ_n^c, i.e, $\{Y_t\}$ is a circular process with period n. We saw in §2.4 that Σ_n^c is a circulant if, and only if, $\Sigma_n^c = F_n \Lambda_n^c \overline{F_n}$, where F_n is the Fourier transform matrix and Λ_n^c is a diagonal matrix whose entries are the DFT of the first row of Σ_n^c. In other words, the singular value decomposition of a circulant can be implemented in $O(n \log n)$ time by taking the FFT of its first row.

If Λ_n^c is p.d. (i.e. Σ_n^c is p.d.) we can obtain a realization Y_0^{n-1} from $N(0, \Sigma_n^c)$ by letting $Y_0^{n-1} = F_n \sqrt{\Lambda_n^c} Z_0^{n-1}$ for $Z_1^{n/2-1} \sim N^c(0, I)$, Z_0, $Z_{n/2}$ iid $N(0,1)$ independently of the other Z's, and $Z_{n-k} = \overline{Z_k}$, where for convenience we assume n is even. Recall that $Z_1^{n/2-1} \sim N^c(0, I)$ means that the real and imaginary parts of the vector $Z_1^{n/2-1}$ are iid $N(0, 1/2)$.

This algorithm can be implemented in $O(n \log n)$ operations: (i) $O(n \log n)$ operations to obtain the FFT of the first row of Σ, (ii) $O(n)$ operations to calculate its square root and multiply by Z_0^{n-1}, and (iii), $O(n \log n)$ operations to obtain the inverse FFT.

In implementing this algorithm we assumed that Σ_n^c is a covariance matrix, i.e., that it is not negative definite. This is equivalent to assuming that the DFT of the first row of Σ_n^c is not negative and thus each of its elements has a real square root.

5.2.2 Simulation of non-circular processes

The algorithm used in §5.2.1 to generate realizations from circular processes in $O(\log n/n)$ operations can also be used to generate realisations from non-circular processes with covariance matrix Σ_n.

The approach in this case is to generate realisations of size $2n$ from a circular process whose covariance matrix is chosen so that any n-long segment of the $2n$-long circular process has covariance matrix Σ_n. We will see below how this is done. This idea is essentially due to Jonas (1984) except that in our implementation we introduce a correction for negative definite matrices.

Consider the $2n$ by $2n$ Toeplits circulant

$$\Sigma_{2n}^c = T(c_0, \ldots, c_{n-1}, 0, c_{n-1}, \ldots, c_1). \tag{5.1}$$

and notice that it has the property that any n by n block originating on the main diagonal is equal to Σ_n.

Recall that since Σ_{2n}^c is a circulant it has the factorization

$$\Sigma_{2n}^c = F_{2n} \Lambda_{2n}^c \overline{F_{2n}}. \tag{5.2}$$

In other words, if Λ_{2n}^c is p.d. any n-long segment of

$$Y_0^{2n-1} = \overline{F_{2n}} {\Lambda_{2n}^c}^{1/2} Z_0^{2n-1}. \tag{5.3}$$

is distributed as $N(0, \Sigma_n)$, where $Z_1^{n-1} \sim N^c(0,1)$, Z_0, Z_n are iid $N(0,1)$ and $Z_k = Z_{2n-k}$, for $1 \le k \le n$.

This idea will only work if Σ_{2n}^c is indeed a covariance matrix, i.e., if it is not negative definite which is equivalent to the entries of Λ_{2n}^c being non-negative. In fact, whether Σ_{2n}^c is p.d. or not is a function of the smoothness of g If g is not smooth enough it may happen that some the entries in Λ_{2n}^c may not be positive. The reason for this is easy to see: Example 2 in §2.4 showed the entries λ_k^c of Λ^c to be the partial sums of the Fourier series of g at the Fourier frequencies k/n, $k = 0, \ldots, 2n-1$, i.e., $\lambda_k = D_{2n+1} * g(k/n)$. The Dirichlet kernel is not positive everywhere and so peaks in the spectrum may be weighted negatively and leak into regions of low spectrum

making some λ_k^c's negative. On the other hand, if $g \in \text{Lip}(1)$, Lemma 2.20 ensures that there exists some N such that for every $m \geq N$, $D_{2m+1} * g(\lambda) > 0$ for all λ.

In our applications $g = \hat{f}_n = \exp(\hat{h}_n) > 0$ where $\hat{h}_n$ is Wahba's estimate of the log spectrum. Thus, given $g > 0$, we can find Σ_{2m}^c p.d. by determining the smallest $m \geq n$ such that the eigenvalues $D_{2m+1} * g(k/m) > 0$ for $k = 0, \ldots, m$.

In practice we only encountered negative-definite Σ_{2n} when simulating short realizations of proceses with highly peaked spectra, e.g., autoregressions with roots near the unit circle. In order to avoid too many increments of size 1 in searching for p.d. Σ_{2m}^c we adopted the strategy of doubling m until the problem disappeared. In other words, initially we simulated series of length $m = 2^k$ where the actual sample size of interest is $n \leq 2^k$, i.e., we constructed Σ_{2m}^c, for m a power of 2. We then calculate the DFT of the first row of Σ_{2m}^c, i.e., we calculated the eigenvalues of Σ_{2m}^c. If some turned out negative, we would take $m = 2^{k+1}$, and so on, until all eigenvalues were positive.

5.2.3 The algorithm

Recall that for a given g we wish to assess the sampling distribution of an estimator $\hat{\theta}_y$ of θ_g by Monte Carlo. We detail the steps of the algorithm below.

Notice that since we are given a spectrum g and not covariances, step 4 estimates the covariances by approximating the integral in equation (2.2) by a sum.

Specifically, the algorithm consists of the following steps:

1. Get g.
2. Set $d \leftarrow 1$ and $m = 2^k \geq n$.
3. Set $d \leftarrow 2d$.
4. Apply the DFT to $g(k/dm), k = 0, ..., dm - 1$ and obtain $c_0, ..., c_{dm/2-1}$.
5. Calculate the eigenvalues λ_k^c, $k = 0, ..., dm - 1$ of Σ_{2m}^c via an application of the DFT to

$$(c_0, c_1, ..., c_{dm/2-1}, 0, c_{dm/2-1}, ..., c_1.)$$

   If any of the λ_k^c's are not positive go back to step 2.
6. Simulate a realisation $Z = Z_0, ..., Z_{dm-1}$ from $N^c(0, \Lambda_{2m}^c)$ via an application of the FFT to $(\lambda_k^c)^{1/2} U_k$, $k = 0, ..., dm - 1$. The U_k are iid $N^c(0,1)$, $k = 1, ..., dm/2 - 1$, $U_k = \overline{U}_{dm/2-k}$, $k = 1 + dm/2, ..., dm - 1$, and $U_0, U_{dm/2}$ are simply iid $N(0,1)$.
7. Apply the FFT to $Z_0, ..., Z_{dm-1}$ and call the result $Y_0, ..., Y_{dm-1}$. Because $U_k = \overline{U}_{dm/2-k}$, the Y's are real and clearly have distribution $N(0, \Sigma_{dm}^c)$.
8. Pick s such that $0 \leq s \leq dm - n$, and thus $Y_0^{n-1} = Y_s, ..., Y_{n-1+s}$ is a realisation from $N(0, \Sigma_n) = N(0, g)$.
9. Calculate $\hat{\theta}_y$.
10. Go back to step 6.

After repeating steps (6)-(10) a large number of times (100-1000) we usually have a reasonable estimate of the sampling distribution of $\hat{\theta}_y$ under g.

Chapter 6

Applications

In this chapter we discuss an application to estimating the sampling distribution of the MLE (assuming only stationarity) of the first correlation coefficient. Specifically, we use the resampling algorithm described in chapter 5 to estimate the bias and variance of this estimator. We then compare the estimates obtained by resampling with Monte-Carlo approximations of the true finite-sample bias and variance, and with asymptotic theory approximations.

It is important to note that the asymptotic approximations require knowledge of the model and for this reason: (i) will not be very useful in practice and (ii) are prone to be more accurate than estimates which could be actually used in practice; they are presented for comparison purposes. On the other hand, the estimates obtained by resampling do not require specific model assumptions beyond stationarity, Gaussianity, and spectral density in Lip(1).

We present results from simulations under two different models, the AR(1) and MA(1) for several parameter values and sample sizes.

6.1 Estimating correlation in the AR(1) and MA(1) models

Suppose that we observe a realization $X_0^{n-1} \sim N(0, f)$ and we wish to assess the sampling distribution of the estimator

$$\hat{\rho} = \frac{\int \exp(i2\pi\omega) I_n^x(\omega) d\omega}{\int I_n^x(\omega) d\omega} = \frac{\sum_t X_t X_{t-1}}{\sum_t X_t^2} \tag{6.1}$$

of

$$\rho = \frac{\int \exp(i2\pi\omega) f(\omega) d\omega}{\int f(\omega) d\omega} = \frac{\int \cos(\omega) f(\omega) d\omega}{\int f(\omega) d\omega}.$$

under two different models: AR(1) and MA(1).

6.1.1 AR(1) Model

The AR(1) model is

$$X_t + \phi X_{t-1} = \epsilon_t$$

where $\{\epsilon_t\}$ are iid $N(0, \sigma^2)$ variates. For this model $\rho = \phi$ and (cf. e.g., Kendall and Stuart, 1976) the variance of the limiting distribution of $\sqrt{n}(\hat{\rho} - \rho)$ is

$$V = (1 - \rho^2)$$

while the bias of $\hat{\rho}$ is given by

$$B \simeq -\frac{1+3\rho}{n-1}.$$

6.1.2 MA(1) Model

The MA(1) model is

$$X_t = \epsilon_t + \psi\epsilon_{t-1}$$

where, again, $\{\epsilon_t\}$ are iid $N(0,\sigma^2)$ variates. For this model $\rho = \psi/(1+\psi^2)$ and (cf. e.g., Kendall and Stuart, 1976) the variance of the limiting distribution of $\sqrt{n}(\hat{\rho}-\rho)$ is

$$V = (1 - 3\rho^2 + 4\rho^4)$$

while the bias of $\hat{\rho}$ is given by

$$B \simeq \frac{1}{n-1}(1+\rho)(4\rho^2 - 2\rho - 1).$$

6.2 Results from simulations

Below we present results from simulations. The object is to estimate bias and variance for the estimator of the first correlation given by (6.1). We compare three quantities:

1. asymptotic theory approximations
2. Monte-Carlo approximations
3. bootstrap estimates

We simulated 6 different processes: four AR(1) processes with parameters .9, -.9, .5, -.5, and two MA(1) processes with parameter -1 and 1. For each of these models we considered three sample sizes: 32, 128, and 512.

6.2.1 Asymptotic approximations

The asymptotic approximations were calculated using the formulæ in the previous section. We report these under the heading ∞.

6.2.2 Monte-Carlo approximations

Monte-Carlo approximations of bias and variability were obtained for each model and each sample size by simulating $N_1 = 10,000$ realizations of the process and reporting the average bias and average square deviation from the mean as the Monte-Carlo estimates of bias and variability. In other words, if $\hat{\rho}_i$ denotes the estimate of the value of the first correlation ρ based on the i^{th} sample, we estimate bias by

$$b_{MC}(\hat{\rho}) = \frac{1}{N_1}\sum_{i=1}^{N_1}\hat{\rho}_i - \rho = \bar{\hat{\rho}} - \rho.$$

Similarly, to estimate variance we used

$$\mathrm{Var}_{MC}(\hat{\rho}) = \frac{1}{N_1}\sum_{i=1}^{N_1}(\hat{\rho}_i - \bar{\hat{\rho}})^2.$$

We report these estimates with estimates of their standard errors in parentheses under the heading $\overline{\mathrm{MC}}$.

6.2.3 Resampling estimates

Resampling estimates were obtained in the following manner: $N_1 = 200$ realization of the original process were simulated and for each an estimate of the spectrum was obtained using Wahba's estimate. In addition, for each of these estimates of the spectrum the associated first correlation was calculated, i.e., if $\hat{f}_n^i$ is the estimate of the spectrum for the i^{th} realization, then

$$\rho_i^* = \frac{\int \cos(2\pi\omega)\hat{f}_n^i(\omega)d\omega}{\int \hat{f}_n^i(\omega)d\omega}$$

is the correlation associated with $\hat{f}_i$.

Then, for each realization (i.e., for each $\hat{f}_n^i$), $N_2 = 200$ realizations from $N(0, \hat{f}_n)$ were simulated and we estimated bias for each by

$$b_{MC}(\hat{\rho}_i^*) = \frac{1}{N_2}\sum_j \hat{\rho}_{ij}^* - \rho_i^* = \overline{\hat{\rho}_i^*} - \rho_i^*$$

and variance by

$$\mathrm{Var}_{MC}(\hat{\rho}_i^*) = \frac{1}{N_2}\sum_j (\hat{\rho}_{ij}^* - \overline{\hat{\rho}_i^*})^2$$

where $\hat{\rho}_{ij}^*$ is the estimate of correlation from the j^{th} bootstrap sample of the i^{th} realization.

Under the heading $\overline{\mathrm{BS}}$ we report the averages of these quantities over the 200 realizations with estimates of their standard errors in parentheses.

6.2.4 Tables

In the following tables we use the notation $1.234\__2$ to mean 1.234×10^{-2}.

$n = 32$	BIAS	VAR
∞	1.613_{-2}	2.344_{-2}
$\overline{\text{MC}}$	$4.443_{-2} \pm (0.152_{-2})$	$2.316_{-2} \pm (0.042_{-2})$
$\overline{\text{BS}}$	$3.392_{-2} \pm (0.162_{-2})$	$2.612_{-2} \pm (0.069_{-2})$
$n = 128$		
∞	0.394_{-2}	0.586_{-2}
$\overline{\text{MC}}$	$1.212_{-2} \pm (0.076_{-2})$	$0.584_{-2} \pm (0.010_{-2})$
$\overline{\text{BS}}$	$1.070_{-2} \pm (0.048_{-2})$	$0.658_{-2} \pm (0.015_{-2})$
$n = 512$		
∞	0.098_{-2}	0.146_{-2}
$\overline{\text{MC}}$	$0.302_{-2} \pm (0.038_{-2})$	$0.142_{-2} \pm (0.0025_{-2})$
$\overline{\text{BS}}$	$0.307_{-2} \pm (0.021_{-2})$	$0.156_{-2} \pm (0.0024_{-2})$

Table 6.1: $X_t + 0.5X_{t-1} = \epsilon_t \Rightarrow \rho_1 = -.5$

$n = 32$	BIAS	VAR
∞	1.613_{-2}	1.563_{-2}
$\overline{\text{MC}}$	$3.054_{-2} \pm (0.124_{-2})$	$1.525_{-2} \pm (0.027_{-2})$
$\overline{\text{BS}}$	$3.291_{-2} \pm (0.116_{-2})$	$1.763_{-2} \pm (0.048_{-2})$
$n = 128$		
∞	0.394_{-2}	0.391_{-2}
$\overline{\text{MC}}$	$0.778_{-2} \pm (0.062_{-2})$	$0.385_{-2} \pm (0.007_{-2})$
$\overline{\text{BS}}$	$0.816_{-2} \pm (0.039_{-2})$	$0.432_{-2} \pm (0.009_{-2})$
$n = 512$		
∞	0.098_{-2}	0.098_{-2}
$\overline{\text{MC}}$	$0.178_{-2} \pm (0.031_{-2})$	$0.094_{-2} \pm (.0017_{-2})$
$\overline{\text{BS}}$	$0.213_{-2} \pm (0.017_{-2})$	$0.109_{-2} \pm (.0016_{-2})$

Table 6.2: $X_t = \epsilon_t - 1.0\epsilon_{t_1} \Rightarrow \rho_1 = -.5$

$n = 32$	BIAS	VAR
∞	-8.065_{-2}	2.344_{-2}
$\overline{\text{MC}}$	$-4.291_{-2} \pm (0.152_{-2})$	$2.319_{-2} \pm (0.043_{-2})$
$\overline{\text{BS}}$	$-3.520_{-2} \pm (0.157_{-2})$	$2.552_{-2} \pm (0.072_{-2})$
$n = 128$		
∞	-1.969_{-2}	0.586_{-2}
$\overline{\text{MC}}$	$-1.077_{-2} \pm (0.077_{-2})$	$0.591_{-2} \pm (0.010_{-2})$
$\overline{\text{BS}}$	$-1.211_{-2} \pm (0.053_{-2})$	$0.661_{-2} \pm (0.016_{-2})$
$n = 512$		
∞	-0.489_{-2}	0.146_{-2}
$\overline{\text{MC}}$	$-0.236_{-2} \pm (0.039_{-2})$	$0.149_{-2} \pm (.0026_{-2})$
$\overline{\text{BS}}$	$-0.267_{-2} \pm (0.021_{-2})$	$0.157_{-2} \pm (.0023_{-2})$

Table 6.3: $X_t - .5X_{t-1} = \epsilon_t \Rightarrow \rho_1 = 0.5$

$n = 32$	BIAS	VAR
∞	-4.839_{-2}	1.563_{-2}
$\overline{\text{MC}}$	$-2.960_{-2} \pm (0.123_{-2})$	$1.520_{-2} \pm (0.027_{-2})$
$\overline{\text{BS}}$	$-3.435_{-2} \pm (0.112_{-2})$	$1.701_{-2} \pm (0.046_{-2})$
$n = 128$		
∞	-1.181_{-2}	0.391_{-2}
$\overline{\text{MC}}$	$-0.734_{-2} \pm (0.063_{-2})$	$0.396_{-2} \pm (0.007_{-2})$
$\overline{\text{BS}}$	$-0.892_{-2} \pm (0.041_{-2})$	$0.461_{-2} \pm (0.011_{-2})$
$n = 512$		
∞	-0.294_{-2}	0.098_{-2}
$\overline{\text{MC}}$	$-0.148_{-2} \pm (0.032_{-2})$	$0.101_{-2} \pm (.0017_{-2})$
$\overline{\text{BS}}$	$-0.190_{-2} \pm (0.018_{-2})$	$0.111_{-2} \pm (.0016_{-2})$

Table 6.4: $X_t = \epsilon_t + 1.0\epsilon_{t_1} \Rightarrow \rho_1 = 0.5$

$n = 32$	BIAS	VAR
∞	5.484_{-2}	0.594_{-2}
$\overline{\text{MC}}$	$7.720_{-2} \pm (0.102_{-2})$	$1.039_{-2} \pm (0.025_{-2})$
$\overline{\text{BS}}$	$6.261_{-2} \pm (0.123_{-2})$	$1.432_{-2} \pm (0.063_{-2})$
$n = 128$		
∞	1.339_{-2}	0.148_{-2}
$\overline{\text{MC}}$	$2.130_{-2} \pm (0.043_{-2})$	$0.187_{-2} \pm (0.004_{-2})$
$\overline{\text{BS}}$	$1.932_{-2} \pm (0.040_{-2})$	$0.243_{-2} \pm (0.009_{-2})$
$n = 512$		
∞	0.333_{-2}	0.037_{-2}
$\overline{\text{MC}}$	$0.523_{-2} \pm (0.020_{-2})$	$0.039_{-2} \pm (.0007_{-2})$
$\overline{\text{BS}}$	$0.517_{-2} \pm (0.014_{-2})$	$0.046_{-2} \pm (.0011_{-2})$

Table 6.5: $X_t + 0.9X_{t-1} = \epsilon_t \Rightarrow \rho_1 = -.9$

$n = 32$	BIAS	VAR
∞	-11.94_{-2}	0.594_{-2}
$\overline{\text{MC}}$	$-7.638_{-2} \pm (0.101_{-2})$	$1.018_{-2} \pm (0.025_{-2})$
$\overline{\text{BS}}$	$-6.302_{-2} \pm (0.137_{-2})$	$1.427_{-2} \pm (0.060_{-2})$
$n = 128$		
∞	-2.913_{-2}	0.148_{-2}
$\overline{\text{MC}}$	$-2.041_{-2} \pm (0.043_{-2})$	$0.187_{-2} \pm (0.004_{-2})$
$\overline{\text{BS}}$	$-1.915_{-2} \pm (0.040_{-2})$	$0.222_{-2} \pm (0.007_{-2})$
$n = 512$		
∞	-0.724_{-2}	0.037_{-2}
$\overline{\text{MC}}$	$-0.504_{-2} \pm (0.020_{-2})$	$0.040_{-2} \pm (.0008_{-2})$
$\overline{\text{BS}}$	$-0.504_{-2} \pm (0.012_{-2})$	$0.043_{-2} \pm (.0010_{-2})$

Table 6.6: $X_t - .9X_{t-1} = \epsilon_t \Rightarrow \rho_1 = 0.9$

Chapter 7

Improving estimators by resampling

7.1 Sufficiency and improved estimators

Suppose that we observe X_0^{n-1} with distribution P_Θ and suppose there exists a sufficient statistic $S_n = S(X_0^{n-1})$ for Θ. In this case, whenever $\hat{\theta}$ is used to estimate θ, $\hat{\theta}^* = E_\Theta(\hat{\theta}|S_n)$ may also be used to estimate θ. Notice that sufficiency is only required to insure that $\hat{\theta}^*$ does not depend on Θ and is hence an estimator.

Now, the Rao-Blackwell theorem states that if the *loss function* $L(\hat{\theta}, \theta) = d(\hat{\theta} - \theta)$ is convex, then the *risk function*

$$R(\hat{\theta}^*, \Theta) = E_\Theta L(\hat{\theta}^*, \theta) \leq R(\hat{\theta}, \Theta)$$

for all Θ. In other words, if there exists a sufficient statistic S_n, $\hat{\theta}^*$ is at least as good an estimator as $\hat{\theta}$ and possibly better.

For example, if the loss is quadratic, i.e., $L(\hat{\theta} - \theta) = (\hat{\theta} - \theta)^2$ the risk is the mean square error:

$$E_\Theta(\hat{\theta} - \theta)^2 = \mathrm{Var}(\hat{\theta}) + \mathrm{b}(\hat{\theta})^2$$

and hence

$$R(\hat{\theta}^*, \Theta) = E_\Theta(E_\Theta(\hat{\theta}|S_n) - \theta)^2 = R(\hat{\theta}, \Theta) - E_\Theta(\mathrm{Var}(\hat{\theta}|S_n)).$$

so that the reduction in risk achieved by using $\hat{\theta}^*$ instead of $\hat{\theta}$ is

$$E_\Theta(\mathrm{Var}(\hat{\theta}|S_n)) \geq 0. \tag{7.1}$$

A simple argument (cf. e.g. Lehmann, 1983) shows that for any estimator $\hat{\theta}(X_0^{n-1})$ there exists a (possible randomised) estimator $\hat{\theta}^r$ based on S_n with the same risk function. Indeed, if S_n is sufficient, the conditional distribution of X_0^{n-1} given S_n does not depend on Θ. This implies that if Y_0^{n-1} is drawn from the conditional distribution of X_0^{n-1} given S_n, the risk function of $\hat{\theta}^r = \hat{\theta}(Y_0^{n-1})$ is the same as that of $\hat{\theta}(X_0^{n-1})$.

The estimator $\hat{\theta}^r$ described above will be a randomised estimator unless it is equal to $\hat{\theta}$ with probability 1. In this case, a corollary to the Rao-Blackwell Theorem asserts that $\hat{\theta}^*$ is uniformly better than $\hat{\theta}^r$ (and hence than $\hat{\theta}$).

The last paragraph motivates the approach we will follow. The heuristic idea is to find a sufficient statistic S_n of lower dimension than the data with the hope that $\hat{\theta}^r$ (and hence $\hat{\theta}$) has a non-degenerate distribution given S_n (and hence a variance) so that $\hat{\theta}^*$ is indeed a better estimator than $\hat{\theta}$ (compare (7.1)).

7.2 Sufficiency in time series: the circular case

If the data are a realization from a stationary Gaussian process, i.e, $X_0^{n-1} \sim N(0, \Sigma_n)$ for Toeplitz Σ_n, there is, in general, no sufficient statistic of dimension lower than n unless we assume the data are generated by a specific model with a reduced number of parameters.

The types of models commonly used for time series data assume parametric representations for either, (i) the mechanism generating the data, e.g., ARMA(p,q), or (ii), for the spectrum. The spectrum in general does not determine the model, but it does under the additional assumptions of Gaussianity and stationarity.

Our strategy for finding an improved estimator $\hat{\theta}^*$ on the basis of $\hat{\theta}$ is, in principle, very simple: apply the Rao-Blackwell Theorem, i.e., find a sufficient statistic S_n under the assumed model for the data and calculate the expectation of $\hat{\theta}$ conditional on S_n. The problem is that under most of these models, even in the Gaussian case, it is not easy to determine a reasonable sufficient statistic, let alone to find the conditional distribution of the data given the sufficient statistic.

Because of the difficulties just mentioned we will take a different approach: we will exploit the fact that the problem simplifies enormously under the assumption that the process is circular. Under this assumption: (i) the periodogram is a sufficient statistic and (ii), the distribution of the data conditional on the periodogram is easy to obtain.

7.2.1 Approximating estimators by Monte-Carlo

Assume we observe X_0^{n-1}. The DFT of the data is $J_n^x = F_n X_0^{n-1} = \Lambda_n^{1/2} \Gamma_0^{n-1}$ where Λ_n is the diagonal matrix of amplitudes of J_n^x and

$$\Gamma_0^{n-1} = \begin{pmatrix} \exp(i2\pi\gamma_0) \\ \vdots \\ \exp(i2\pi\gamma_{n-1}) \end{pmatrix} \tag{7.2}$$

is the vector of phases. Since X_0^{n-1} is real we have that $\lambda_j = \lambda_{n-j}$ and $\gamma_j = -\gamma_{n-j}$.

Under the assumption of circularity, γ_j are iid $U(0,1)$ for $j = 1, \ldots, n/2 - 1$. This means that, given the periodogram, the amplitude of the DFT is fixed and all we have to do to obtain a sample Y_0^{n-1} with the same conditional (on the periodogram) distribution as the original data X_0^{n-1} is to draw iid $U(0,1)$ phases Δ_0^{n-1}, multiply each $\exp(i2\pi\delta_j)$ by the corresponding amplitude $\lambda_j^{1/2}$ (given by the square root of the periodogram), and finally take the inverse DFT. Using this device, many new samples with the same conditional distribution as the original data may be generated and, for each, the value of $\hat{\theta}$ can be calculated and averaged with the rest to yield an arbitrarily precise approximation $\hat{\theta}^*$ to $E_\Theta(\hat{\theta}|S_n)$.

The obvious question which arises is: why go to all this trouble? Why not simply start off with an efficient estimator (e.g., the MLE) and avoid (usually expensive) resampling? The answer is that efficient estimators may not always be known, or may not have simple representations, but there may be other estimators, albeit inefficient, which are easy to calculate. This point is illustrated with an example.

7.2.2 Example: crossing rates

Suppose we are interested in estimating the rate ϕ_α at which a process crosses a barrier at height α, i.e., the rate at which the process $= X_k - \alpha$ changes sign. In other words, if we define

$$Z_k = \begin{cases} 1 & \text{if } X_k \geq \alpha \\ 0 & \text{otherwise,} \end{cases} \tag{7.3}$$

we wish to assess

$$\phi_\alpha = \lim_{n\to\infty} \frac{1}{n-1} \sum_{k=0}^{n-2} (Z_{k+1} - Z_k)^2.$$

Since $(Z_{k+1} - Z_k)^2 = 1$ if, and only if, a crossing of α ocurrs at time $k+1$, the problem reduces to evaluating

$$\begin{aligned} \phi_\alpha &= E(Z_{k+1} - Z_k)^2 \\ &= EZ_{k+1}^2 + EZ_k^2 - 2EZ_{k+1}Z_k \\ &= 2P(X_k \geq \alpha) - 2P(X_{k+1} \geq \alpha \cap X_k \geq \alpha) \\ &= 2\Phi(\alpha/\sigma_x) - 2P(X_{k+1} \geq \alpha \cap X_k \geq \alpha) \end{aligned}$$

Even in the Gaussian case there is no simple closed form representation for $\hat{\phi}_\alpha$. However, in the special case when $\alpha = 0$, we have that (cf. e.g. Johnson & Kotz, 1972)

$$2P(X_{k+1} \geq 0 \cap X_k \geq 0) = .5 + \pi^{-1} \arcsin(\rho)$$

where $\rho = EX_0X_1/EX_0^2$, so that

$$\phi_0 = .5 - \pi^{-1} \arcsin(\rho).$$

An efficient estimator of ρ is the MLE (under the Toeplitz and the AR(1) models) $\hat{\rho} = \sum X_k X_{k+1} / \sum X_k^2$ so that $\hat{\phi}_0 = .5 - \pi^{-1} \arcsin(\hat{\rho})$ is the MLE of ϕ_0.

On the other hand, consider the naive estimator of ϕ_α given by

$$\hat{\phi}_\alpha^o = \frac{1}{n-1} \sum_{0}^{n-2} (Z_{k+1} - Z_k)^2.$$

Notice that $\hat{\phi}_\alpha^o$ can be calculated easily for all α but it is clearly inefficient since it uses only the information in the sign of $\{X_k - \alpha\}$.

As we mentioned above, under the assumption that our process is circular we can resample the phases of the DFT of our original realization to construct an arbitrarily close approximation to a third estimator of ϕ_α,

$$\hat{\phi}_\alpha^* = E_\Theta \left(\hat{\phi}^o | I_n^x \right)$$

where I_n^x is the periodogram of the data X_0^{n-1} evaluated at the Fourier frequencies.

Two questions arise now: first, what are the relative efficiencies of these three estimators and, second, in what situations, if any, can we improve our estimators by resampling if the original process is not circular? We will defer addressing the first question until the simulations and try to shed some light on the second one.

7.3 Improved estimators: non-circular case

Our approach is suggested by that of the circular case: calculate the periodogram of the original data and estimate the random variable $E(\hat{\theta}|I_n^x)$ by $\hat{\theta}^*$ where $\hat{\theta}^*$ is obtained by resampling our original data conditional on the periodogram and pretending that the phases are iid $U(0,1)$ to obtain new samples. For each new sample we then calculate the value of $\hat{\theta}$ and, finally, take the mean of the values of $\hat{\theta}$ for each sample to be our approximation to $\hat{\theta}^*$.

Since in the non-circular case the phases are not really iid $U(0,1)$, but only approximately so, we will have that $\hat{\theta}^* \neq E\{\hat{\theta}|I_n^x\}$ in general (the r.h.s. may not even be a statistic since it may depend on θ), but if the difference is not large we may still get an improvement over $\hat{\theta}$. For example, if $\hat{\theta}$ is unbiased, then $\hat{\theta}^*$ will in general be biased but will have lower variance so that in the end the MSE may be reduced by using $\hat{\theta}^*$.

7.3.1 Distribution of resampled data

In order to study more precisely what effect pretending the phases are iid $U(0,1)$ has on the resampled data Y_0^{n-1}, we determine the distribution of Y_0^{n-1}.

Lemma 7.1 *Let* $X_0^{n-1} \sim N(0, \Sigma_n)$, $\Sigma_n = T(c_0, \ldots, c_{n-1})$, *and* $J_n^x = F_n X_0^{n-1} = \Lambda_n^{1/2} \Gamma_0^{n-1}$ *where* Λ_n *is the diagonal matrix of amplitudes* $I_n^x(k/2n)$, $k = 0, \ldots, n-1$. *If*

$$\Delta_0^{n-1} = \begin{pmatrix} \exp(i2\pi\delta_0) \\ \vdots \\ \exp(i2\pi\delta_{n-1}) \end{pmatrix}$$

where the components of $\Delta_1^{n/2-1}$ *are iid* $U(0,1)$, $\delta_k = \delta_{n-k}$ *for* $0 \le k \le n$, *and* $Y_0^{n-1} = \overline{F_n} \Lambda_n^{1/2} \Delta_0^{n-1}$, *then*

$$\mathcal{L}(Y_0^{n-1} | I_n^x) = \mathcal{L}(\tilde{X}_0^{n-1} | I_n^{\tilde{x}} = I_n^x)$$

where $\tilde{X}_0^{n-1} \sim N(0, \Sigma_n^c)$, $\Sigma_n^c = T(c_0^c, \ldots, c_{n-1}^c)$, *and* $c_k^c = (1 - |k|/n)c_k + (|k|/n)c_{n-|k|}$ *for* $0 \le k \le n-1$.

Remark: Lemma 7.1 establishes that, for n large, the covariances c_k^c of the data resampled, pretending the process to be circular, approximate those of the original data for k not too large.

7.3.2 Heuristics

We can now study what $\hat{\theta}^*$ estimates in the non-circular case.

Assume that the parameter of interest is a function of the unkown covariance structure $(c_0, c_1, \ldots)$ of the process, i.e., $\theta_{\Sigma_n} = \theta(f)$. The corresponding parameter for the circular process is $\theta_{f_n} = \theta(f_n)$ where f_n denotes the following "sum" of the first n Fourier coefficients of f, i.e.,

$$f_n(\omega) = \sum_{k=-n+1}^{n-1} ((1 - |k|/n)c_k + (|k|/n)c_{n-|k|}) \exp(i2\pi\omega k).$$

Assume, for simplicity, that $\hat{\theta}$ is unbiased, i.e., $E(\hat{\theta}) = \theta_f$, so that its MSE is just its variance, i.e.,

$$E(\hat{\theta} - \theta_f)^2 = \mathrm{Var}(\hat{\theta}).$$

On the other hand the distribution of the resampled data is $N(0, f_n)$ and since $\hat{\theta}$ is unbiased it follows that $E(\hat{\theta}^*) = \theta_{f_n}$, and hence the MSE of $\hat{\theta}^*$ is

$$E(\hat{\theta}^* - \theta_f)^2 = \mathrm{Var}(\hat{\theta}^*) + (\theta_{f_n} - \theta_f)^2.$$

It remains to assess $b_n \equiv \theta_{f_n} - \theta_f$. To do this it will be convenient to consider θ as a function of the Fourier coefficients of f, i.e.,

$$\theta(f) = \theta(c_0, c_1, \ldots).$$

Similarly

$$\theta(f_n) = \theta(c_0, (1 - 1/n)c_1 + c_{n-1}/n, \ldots, (1 - 1/n)c_1 + c_{n-1}/n, 0, \ldots),$$

so that, using a Taylor expansion, it is easy to see that

$$\begin{aligned} b_n &= \theta_{f_n} - \theta_f \\ &\simeq -\frac{1}{n} \sum_{k=0}^{n-1} k(c_k - c_{n-k}) \frac{\partial \theta}{\partial c_k} - \sum_{|k| \ge n} c_k \frac{\partial \theta}{\partial c_k} \end{aligned}$$

The quantity b_n measures the change that circularising by assuming iid random phases induces in the parameter of interest. If $|b_n|$ is not too large and $\hat{\theta}$ is rather inefficient, we may do substantially better by using $\hat{\theta}^*$ rather than $\hat{\theta}$ even in the non-circular case.

In the example concerning the estimation of the crossing rate of 0, $\phi_0 = .5 - \pi^{-1}\sin^{-1}(\rho)$, $\rho = c_0/c_1$, and it is straightforward to verify that

$$b_n \simeq -\frac{\rho}{n\pi\sqrt{1-\rho^2}}$$

and the results from our simulations indicate that there is considerable reduction of MSE. In fact, it turns out that $\hat{\phi}_0^*$ is almost as efficient as the MLE even in cases where the parameter space is of much smaller dimension than the periodogram. The reason for this is that the bias introduced by assuming circularity is of the same order of magnitude as the bias of the MLE and thus contributes little to the MSE for n large.

7.4 Results from simulations

In this section we report results from the simulations. For each process we simulated $N_1 = 1,000$ realisations and estimated phi_0 by $\hat{\phi}_0^o$, $\hat{\phi}_0$, and $\hat{\phi}_0^*$, reported under the headings naive, MLE, and resampled, respectively. The resampled estimator, $\hat{\phi}_0^*$, was obtained by resampling the phases of each realisations 100 times, i.e., it is the average of 100 $\hat{\phi}_0^o$'s.

$n = 512$	BIAS	MSE
Naive	-0	$0.520_{-3} \pm (.030_{-3})$
MLE	$-0.360_{-2} \pm (.048_{-2})$	$0.225_{-3} \pm (.013_{-3})$
Resampled	$-0.184_{-2} \pm (.047_{-2})$	$0.225_{-3} \pm (.013_{-3})$
b_n	-0.128_{-2}	
$n = 128$		
Naive	-0	$2.022_{-3} \pm (.114_{-3})$
MLE	$-1.141_{-2} \pm (.098_{-2})$	$0.962_{-3} \pm (.063_{-3})$
Resampled	$-0.464_{-2} \pm (.099_{-2})$	$0.975_{-3} \pm (.061_{-3})$
b_n	-0.513_{-2}	

Table 7.1: $X_t + .9X_{t-1} = \epsilon_t \Rightarrow \phi_0 = .8564$

$n = 128$	BIAS	MSE
Naive	0	$2.093_{-3} \pm (.112_{-3})$
MLE	$1.066_{-2} \pm (.095_{-2})$	$0.893_{-3} \pm (.055_{-3})$
Resampled	$0.349_{-2} \pm (.097_{-2})$	$0.945_{-3} \pm (.056_{-3})$
b_n	0.513_{-2}	

Table 7.2: $X_t - .9X_{t-1} = \epsilon_t \Rightarrow \phi_0 = .1436$

Appendix A

Auxiliary results and proofs

A.1 Chapter 2

Proof of Lemma 2.1: Fix n and let $\Sigma_n = T(c_0, \ldots, c_{n-1})$. For any vector $(x_0, \ldots, x_{n-1})$ define $\phi_x(\lambda) = \sum_{s=0}^{n-1} x_s \exp(-i2\pi s\lambda)$. Then, since $f > 0$ and ϕ is continuous,

$$0 < \int_0^1 |\phi_x(\lambda)|^2 f(\lambda) d\lambda = \sum_{s,t=0}^{n-1} c_{s-t} x_s x_t = x' \Sigma_n x$$

unless $|\phi_x(\lambda)| \equiv 0 \Leftrightarrow x_s \equiv 0$. ■

Proof of lemma 2.2

$$|f^{(k)}(\omega)| = |\sum c_s (i2\pi s)^k \exp(i2\pi s\omega)| \leq (2\pi)^k \sum |s|^k |c_s|$$

which proves the first implication of the first part of the lemma. As for the second implication, it is easy to see that if a function f has a bounded first derivative then $M_f = \sup_x |f'(x)|$ and hence $f \in \text{Lip}(1)$.

To show f continuous: Let $\epsilon > 0$. Since $\sum |c_s| < \infty$ there exists $N = N(\epsilon)$ such that $\sum_{|s|>N} |c_s| < \epsilon/2$, so that

$$\begin{aligned} |f(\omega+\delta) - f(\omega)| &= \left| \sum_s c_s (1 - \exp(i2\pi s\delta)) \exp(i2\pi s\omega) \right| \\ &\leq 2 \sum_{|s| \leq N} |c_s \sin(\pi s\delta)| + \epsilon \\ &\leq 2\pi\delta \sum_{s \leq N} |c_s||s| + \epsilon \end{aligned}$$

but for fixed N, the first term above is continuous in δ and goes to 0 as $\delta \to 0$. ■

Proof of lemma 2.4: The first assertion follows immediately from the properties of $Z(\omega)$. As for the second, notice

$$|F - G|^2 = f + g - 2\sqrt{fg}\cos(\theta - \phi)$$

where θ and ϕ are the phases of F and G respectively. Clearly $\theta = \phi$ achieves the minimum. ■

Proof of Lemma 2.5: Obvious. ■

Proof of Lemma 2.10: The first assertion is standard (cf. e.g. Hasminskii and Ibragimov, 1986). As for the second, let

$$\begin{aligned} A_n &= \left|\int \{f(x)-f(s)\}D_n(x-t)D_n(x-s)dx\right| \\ &= \left|\int \{f(x+t)-f(s)\}D_n(x+t-s)D_n(x)dx\right| \\ &\leq \int |f(x+t)-f(s)||D_n(x+t-s)||D_n(x)|dx \\ &\leq M_f \int \min(|x+t-s|,|1-x-t+s|)|D_n(x+t-s)||D_n(x)|dx \end{aligned}$$

since $f \in \text{Lip}(1)$, periodic and symmetric, so that

$$A_n \leq M_f \sup_{0\leq x\leq .5} |xD_n(x)| \int |D_n(x)|dx.$$

Since $\int |D_n(x)|dx = O(\log n)$ (cf. e.g., Butzer and Nessel, 1971), to establish the lemma it suffices to show that there exists K such that $\sup_{0\leq x\leq .5}|xD_n(x)| < K$. Indeed,

$$\begin{aligned} \sup_{0\leq x\leq .5} |xD_n(x)| &= \sup_{0\leq x\leq .5} \left|\frac{x\sin(\pi n x)}{\sin(\pi x)}\right| \leq \\ &\leq \sup_{0\leq x\leq .5} \left|\frac{x}{\sin(\pi x)}\right| \leq \frac{3}{5\pi} \end{aligned}$$

since (cf. Butzer and Nessel, 1971),

$$\frac{\sin(\pi x)}{\pi x} \geq \frac{1-x^2}{1+x^2}.$$

and the lemma follows immediately. ■

Proof of lemma 2.11: The first assertion is obvious. The second is a standard result (cf. e.g. Priestley, 1981) and is also a consequence of the proof of Lemma 2.12 which follows. ■

Proof of lemma 2.12: Let $\lambda = l/n$ and $\mu = m/n$. Using the Cramer representation for X_t and the fact that $\sum_{t=0}^{n-1}\exp(i2\pi\lambda t) = \exp\{i\pi(n-1)\lambda\}D_n(\lambda)$ it is easy to see that

$$J_n^x(\mu) = \frac{1}{\sqrt{n}}\int F(\omega)e^{i\pi(n-1)(\omega-\mu)}D_n(\omega-\mu)dZ(\omega). \tag{A.1}$$

Using the fact that $F(\omega) = \overline{F(-\omega)}$, the symmetry of D_n, and straightforward manipulation, we have that

$$\begin{aligned} \Re\{J_n^x(\mu)\} &= \frac{1}{2}\{J_n^x(\mu) + \overline{J_n^x(\mu)}\} \\ &= \frac{1}{2\sqrt{n}}\int F(\omega)[\exp(i\pi(n-1)(\omega-\mu))D_n(\omega-\mu) \\ &\quad + \exp(i\pi(n-1)(\omega+\mu))D_n(\omega+\mu)]dZ(\omega) \end{aligned}$$

and

$$\begin{aligned} \Im\{J_n^x(\mu)\} &= \frac{1}{2i}\{J_n^x(\mu) - \overline{J_n^x(\mu)}\} \\ &= \frac{1}{2\sqrt{n}}\int F(\omega)[\exp(i\pi(n-1)(\omega-\mu))D_n(\omega-\mu) \\ &\quad - \exp(i\pi(n-1)(\omega+\mu))D_n(\omega+\mu)]dZ(\omega). \end{aligned}$$

Using (2.3) and the symmetry of f and D_n, it is straightforward to see that

$$\begin{aligned}
A_n &= E\Re\{J_n^x(\mu)\}\Re\{J_n^x(\lambda)\} \\
&= \frac{1}{4n}\int f(\omega)\{\exp(i\pi(n-1)(-\mu+\lambda))D_n(\omega-\mu)D_n(\omega-\lambda) \\
&\quad +\exp(i\pi(n-1)(-\mu-\lambda))D_n(\omega-\mu)D_n(\omega+\lambda) \\
&\quad +\exp(i\pi(n-1)(\mu+\lambda))D_n(\omega+\mu)D_n(\omega-\lambda)) \\
&\quad +\exp(i\pi(n-1)(\mu-\lambda))D_n(\omega+\mu)D_n(\omega+\lambda)\}d\omega \\
&= \frac{\cos(\pi(n-1)(\mu-\lambda))}{2n}\int f(\omega)D_n(\omega-\mu)D_n(\omega-\lambda)d\omega \\
&\quad +\frac{\cos(\pi(n-1)(\mu+\lambda))}{2n}\int f(\omega)D_n(\omega+\mu)D_n(\omega-\lambda)d\omega \\
&= \frac{\cos(\pi(\mu-\lambda))}{2n}\int f(\omega)D_n(\omega-\mu)D_n(\omega-\lambda)d\omega \\
&\quad +\frac{\cos(\pi(\mu+\lambda))}{2n}\int f(\omega)D_n(\omega+\mu)D_n(\omega-\lambda)d\omega,
\end{aligned}$$

and the result follows from lemma 2.10.

Similarly,

$$\begin{aligned}
B_n &= E\Im J_n^x(\mu)\Im J_n^x(\lambda) \\
&= \frac{1}{4n}\int f(\omega)\{\exp(i\pi(n-1)(-\mu+\lambda))D_n(\omega-\mu)D_n(\omega-\lambda) \\
&\quad -\exp(i\pi(n-1)(-\mu-\lambda))D_n(\omega-\mu)D_n(\omega+\lambda) \\
&\quad -\exp(i\pi(n-1)(\mu+\lambda))D_n(\omega+\mu)D_n(\omega-\lambda)) \\
&\quad +\exp(i\pi(n-1)(\mu-\lambda))D_n(\omega+\mu)D_n(\omega+\lambda)\}d\omega \\
&= \frac{\cos(\pi(n-1)(\mu-\lambda))}{2n}\int f(\omega)D_n(\omega-\mu)D_n(\omega-\lambda)d\omega \\
&\quad -\frac{\cos(\pi(n-1)(\mu+\lambda))}{2n}\int f(\omega)D_n(\omega+\mu)D_n(\omega-\lambda)d\omega \\
&= \frac{\cos(\pi(\mu-\lambda))}{2n}\int f(\omega)D_n(\omega-\mu)D_n(\omega-\lambda)d\omega \\
&\quad -\frac{\cos(\pi(\mu+\lambda))}{2n}\int f(\omega)D_n(\omega+\mu)D_n(\omega-\lambda)d\omega
\end{aligned}$$

and

$$\begin{aligned}
C_n &= E\Re J_n^x(\mu)\Im J_n^x(\lambda) \\
&= \frac{1}{4n}\int f(\omega)\{\exp(i\pi(n-1)(-\mu+\lambda))D_n(\omega-\mu)D_n(\omega-\lambda) \\
&\quad -\exp(i\pi(n-1)(-\mu-\lambda))D_n(\omega-\mu)D_n(\omega+\lambda) \\
&\quad +\exp(i\pi(n-1)(\mu+\lambda))D_n(\omega+\mu)D_n(\omega-\lambda)) \\
&\quad -\exp(i\pi(n-1)(\mu-\lambda))D_n(\omega+\mu)D_n(\omega+\lambda)\}d\omega \\
&= -\frac{\sin(\pi(n-1)(\mu-\lambda))}{2n}\int f(\omega)D_n(\omega-\mu)D_n(\omega-\lambda)d\omega \\
&\quad +\frac{\sin(\pi(n-1)(\mu+\lambda))}{2n}\int f(\omega)D_n(\omega+\mu)D_n(\omega-\lambda)d\omega
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\sin(\pi(\mu-\lambda))}{2n}\int f(\omega)D_n(\omega-\mu)D_n(\omega-\lambda)d\omega \\
&\quad -\frac{\sin(\pi(\mu+\lambda))}{2n}\int f(\omega)D_n(\omega+\mu)D_n(\omega-\lambda)d\omega \\
&= M_f O(\log n/n).
\end{aligned}
$$

■

Proof of lemma 2.13: To prove the first assertion, let

$$U_1 = \frac{1}{\sqrt{2}}\left\{X\sqrt{1+\rho}+Y\sqrt{1-\rho}\right\}$$

and

$$U_2 = \frac{1}{\sqrt{2}}\left\{X\sqrt{1-\rho}-Y\sqrt{1+\rho}\right\}$$

for X, Y iid $N(0,1)$. It is easy to verify that U_1, U_2 are also iid $N(0,1)$,

$$X = \frac{1}{\sqrt{2}}\left\{U_1\sqrt{1+\rho}+U_2\sqrt{1-\rho}\right\},$$

$$Y = \frac{1}{\sqrt{2}}\left\{U_1\sqrt{1-\rho}-U_2\sqrt{1+\rho}\right\},$$

and $W = (1+\rho)U_1^2 + (1-\rho)U_2^2$. This proves the first assertion.

To prove the second assertion it suffices to show that the random variables $Z_1 = U_1^2 + U_2^2$ and $Z_2 = (U_1^2 - U_2^2)/(U_1^2 + U_2^2)$ are independent.

Indeed, verify $U_2^2 = Z_1(1-Z_2)/2$, $U_1^2 = Z_1(1+Z_2)/2$. Then, verify the Jacobian is $Z_1/2$. At last, verify the resulting joint density of Z_1 and Z_2 is given by

$$\pi^{-1}\exp(-z_1/2)(1-z_2^2)^{-1/2}$$

and the result follows. ■

Proof of lemma 2.14: Using the Cramer representation (2.3), it is straightforward to see that

$$I_n^x(\lambda)I_n^y(\mu) = \frac{1}{n^2}\int\int\int\int F(\omega_0)\overline{F(\omega_0')}G(\omega_1)\overline{G(\omega_1')}e^{i\pi(n-1)(\omega_0-\omega_0'+\omega_1-\omega_1')}\times$$

$$\times D_n(\omega_0-\lambda)D_n(\omega_0'-\lambda)D_n(\omega_1-\mu)D_n(\omega_1'-\mu)dZ(\omega_0)\overline{dZ(\omega_0')}dZ(\omega_1)\overline{dZ(\omega_1')}.$$

Taking expectation yields contributions in three cases:

1. $\omega_0 = \omega_0'$ and $\omega_1 = \omega_1'$
2. $\omega_0 = \omega_1'$ and $\omega_1 = \omega_0'$
3. $\omega_0 = -\omega_1$ and $\omega_0' = -\omega_1'$.

The contribution from (1) is

$$
\begin{aligned}
&\frac{1}{n}\int f(\omega_0)D_n^2(\omega_0-\lambda)d\omega_0\frac{1}{n}\int g(\omega_1)D_n^2(\omega_1-\mu)d\omega_1 \quad = \\
&\qquad \int f(\omega)K_n(\omega-\lambda)d\omega\int g(\omega)K_n(\omega-\lambda)d\omega \quad = \quad EI_n^x(\lambda)EI_n^y(\mu).
\end{aligned}
$$

The contribution from (2) is

$$\frac{1}{n}\int F(\omega_0)\overline{G(\omega_0)}e^{i\pi(n-1)(\mu-\lambda)}D_n(\omega_0-\lambda)D_n(\omega_0-\mu)d\omega_0\times$$

$$\times \frac{1}{n} \int G(\omega_1)\overline{F(\omega_1)} e^{-i\pi(n-1)(\mu-\lambda)} D_n(\omega_1 - \lambda) D_n(\omega_1 - \mu) d\omega_1 =$$

$$= \frac{1}{n^2} \left| \int F(\omega)\overline{G(\omega)} D_n(\omega - \lambda) D_n(\omega - \mu) d\omega \right|^2 .$$

Analogously, the contribution from (3) is

$$\frac{1}{n^2} \left| \int F(\omega)\overline{G(\omega)} D_n(\omega - \lambda) D_n(\omega + \mu) d\omega \right|^2 .$$

The result now follows trivially ■

Proof of lemma 2.16: Consider each of the terms in (2.9). By Lemma 2.10, the first term

$$\int f(x) D_n(x - \lambda) D_n(x - \mu) dx = f(\mu) D_n(\mu - \lambda) + \int \{f(x) - f(\mu)\} D_n(x - \lambda) D_n(x - \mu) dx.$$

and a similar expression holds for the second term. If $\mu \neq \lambda$ are Fourier frequencies, $0 \leq \mu, \lambda \leq .5$, $D_n(\mu - \lambda) = D_n(\mu + \lambda) = 0$ and Lemma 2.10 gives the desired result. ■

Before we can prove lemma 2.17 we need three auxiliary results. First a result bounding the distribution of quadratic forms:

Lemma A.1 *If $X \sim N(\mu, \Sigma), \Sigma > 0$, then there exists a random variable $W \sim \chi^2_n(\mu'\Sigma^{-1}\mu)$ such that*

$$t_* W \leq X'X \leq t^* W$$

where t_*, t^* are the smallest and largest eigenvalues of Σ respectively.

Proof: Let $\Sigma = JTJ'$ where T is the diagonal matrix whose entries are the eigenvalues of Σ and J is the matrix of normalized eigenvectors such that $JJ' = J'J = I$. Notice that $X = JT^{1/2}(T^{-1/2}J'\mu + Y)$ for some $Y \sim N(0, I)$ so that

$$X'X = (Y' + \mu'JT^{-1/2})T(T^{-1/2}J'\mu + Y) = Z'TZ$$

for $Z = T^{-1/2}J'\mu + Y \sim N(T^{-1/2}J'\mu, I)$.

But $Z'TZ = \sum t_i Z_i^2$ so that clearly

$$t_* \sum Z_i^2 \leq Z'TZ \leq t^* \sum Z_i^2$$

and the result follows from the fact that $\sum Z_i^2 \sim \chi^2_n(E(Z)'E(Z))$ and that $E(Z)'E(Z) = \mu'JT^{-1}J'\mu = \mu'\Sigma^{-1}\mu$. ■

Second, a lemma which bounds the expectation of the log of a non-central χ^2_2:

Lemma A.2 *If $W \sim \chi^2_2(2\theta^2)$ is a random variable with non-central χ^2 distribution with 2 degrees of freedom and non-centrality parameter $2\theta^2$, then*

$$-\gamma \leq E \log(W/2) \leq \theta^2(1 + \gamma) - \gamma$$

where $\gamma = .57721\ldots$ is Euler's constant.

Proof: Bickel and Doksum (1977) show that a $\chi^2_2(2\theta^2)$ variate can be generated by first sampling a Poisson (θ^2) variate R and then sampling a central χ^2_{2R+2} variate independently. Also, it is well known that

$$E \log(\chi^2_{2r}/2) = \frac{\Gamma'(r)}{\Gamma(r)} = \psi(r)$$

where ψ is the *digamma* function (cf. e.g. Abramowitz and Stegun, 1954). In particular, $\psi(1) = -\gamma$, so that

$$\begin{aligned} E(\log W/2) &= \sum_{k=0}^{\infty} \frac{\exp(-\theta^2)\theta^{2k}\psi(k+1)}{k!} \\ &= \exp(-\theta^2)(\psi(1) + \sum_{k=1}^{\infty} \frac{\theta^{2k}\psi(k+1)}{k!}) \end{aligned}$$

Since $0 \leq \psi(k+1) \leq k$, for $k \geq 1$ (cf. Abramowitz and Stegun, 1954), all the terms in the sum are positive so that the sum itself is positive which implies

$$\exp(-\theta^2)\psi(1) \leq E \log W/2.$$

Since $\psi(1) = -\gamma$ and $0 \leq \exp(-\theta^2) \leq 1$ the first inequality follows.

The second follows from the same property, $0 \leq \psi(k+1) \leq k$, $k \geq 1$, so that

$$\sum_{k=1}^{\infty} \frac{\theta^{2k}\psi(k+1)}{k!} \leq \sum_{k=1}^{\infty} \frac{\theta^{2k}}{(k-1)!} = \theta^2 \exp(\theta^2).$$

Thus, we have that

$$E \log W/2 \leq -\gamma \exp(-\theta^2) + \theta^2 \leq -\gamma(1-\theta^2) + \theta^2 = -\gamma + \theta^2(1+\gamma). \blacksquare$$

Lemma A.3 *If U and V are two-dimensional vectors such that*

$$\begin{pmatrix} U \\ V \end{pmatrix} \sim N\left(0, \begin{pmatrix} I & \tau A \\ \tau A' & I \end{pmatrix}\right),$$

where τ^2 is the largest eigenvalue of $\tau^2 A'A$ and $h(U)$ is a real-valued random variable with finite second moment σ^2, then $\operatorname{cov}(h(U), \log V'V) = \sigma O(\tau^2)$, uniformly in A and h.

Proof of Lemma A.3:

$$\begin{aligned} &\operatorname{cov}(h(U), \log V'V) = \operatorname{cov}(h(U), \log V'V/2) = \\ &E\{(h(U) - Eh(U))(\log V'V/2 - E \log V'V/2)\} = \\ &E\{(h(U) - Eh(U))E(\log V'V/2 - E \log V'V/2)|U\}. \end{aligned} \tag{A.2}$$

First we evaluate the conditional expectation above. We know $E(\log V'V/2) = -\gamma$ and the distribution of V given U is simply $N(\tau A'U, I - \tau^2 A'A)$. Write $A'A = P\Lambda P'$, for orthogonal P and diagonal Λ with entries 1, λ_2, where $0 \leq \lambda_2 \leq 1$.

First we bound $V'V|U$: apply lemma A.1 with $\Sigma = I - \tau^2 A'A = P(I - \tau^2\Lambda)P'$ and $\mu = \tau A'U = \tau P\Lambda^{1/2}QU$, where Q is orthogonal. If we let $QU = R$, it is easy to verify that the non-centrality parameter is given by

$$\begin{aligned} 2\theta^2 &= \tau^2 U'Q'\Lambda^{1/2}(I - \tau^2\Lambda)^{-1}\Lambda^{1/2}QU \\ &= \tau^2\left(R_1^2 \frac{1}{1-\tau^2} + R_2^2 \frac{\lambda_2}{1-\tau^2\lambda_2}\right) \\ &\leq R'R\frac{\tau^2}{1-\tau^2} = U'U\frac{\tau^2}{1-\tau^2}. \end{aligned}$$

Since the eigenvalues of $I - \tau^2 A'A$ are given by $1 - \tau^2 = t_* \leq t^* = 1 - \tau^2\lambda_2$ it follows that

$$(1-\tau^2)W \leq V'V|U \leq (1-\tau^2\lambda_2)W, \tag{A.3}$$

for some $W \sim \chi^2_2(2\theta^2)$. Taking logs and expectation in A.3, an application of lemma A.2 yields

$$\log(1-\tau^2) \leq E\left(\log\frac{V'V}{2}|U\right) + \gamma \leq \log(1-\tau^2\lambda_2) + \frac{U'U}{2}\frac{\tau^2}{1-\tau^2}(1+\gamma)$$

which implies

$$\left|E\left(\log\frac{V'V}{2}|U\right) + \gamma\right| \leq \log\left(\frac{1}{1-\tau^2}\right) + \frac{U'U}{2}\frac{\tau^2}{1-\tau^2}(1+\gamma) = (1+\gamma)\left(\frac{U'U}{2}+1\right)O(\tau^2).$$

Lemma A.3 follows upon using this result and applying Schwarz's inequality to evaluate (A.2) together with the fact that $E(U'U)^2 = E\chi^4_2 = 8$ and $h(U)$ has standard deviation σ. ■

We can now prove lemma 2.17: By lemma 2.13 the two periodogram ordinates can be written as

$$\begin{aligned} W_1 &= (U_1^2+U_2^2)(1+\rho_1\theta_1) \\ &= W_1^*(1+\rho_1\theta_1) \end{aligned}$$

where $\rho_1 = M_f O(\log n/n)$, and

$$\begin{aligned} W_2 &= (V_1^2+V_2^2)(1+\rho_2\theta_2) \\ &= W_2^*(1+\rho_2\theta_2). \end{aligned}$$

Since $|\theta_i| \leq 1$

$$\begin{aligned} \log(W_i) &= \log(W_i^*) + \log(1+\rho_i\theta_i) \\ &= \log(W_i^*) + \rho_i\theta_i + O(\rho_i^2). \end{aligned}$$

so that,

$$\begin{aligned} \mathrm{cov}(\log(W_1),\log(W_2)) &= \mathrm{cov}(\log(W_1^*),\log(W_2^*)) + \rho_2\mathrm{cov}(\log W_1^*,\theta_2) \\ &\quad +\rho_1\mathrm{cov}(\log W_2^*,\theta_1) + O(|\rho_1|+|\rho_2|)^2. \end{aligned} \tag{A.4}$$

By lemma 2.12 the last term above is $M_f^2 O(\log n/n)^2$.

It remains to calculate each of the covariances in (A.4). Notice that U_1 and U_2 are independent and so are V_1 and V_2, so that the covariance matrix of U_1, U_2, V_1, and V_2 looks like

$$\begin{pmatrix} I & \tau A \\ \tau A' & I \end{pmatrix}.$$

Lemma 2.12 tells us that τ, ρ_1, and ρ_2 are $M_f O(\log n/n)$ and an application of Lemma A.3 yields

$$\mathrm{cov}(\log(W_1),\log(W_2)) = O(\tau^2) = M_f^2 O(\log n/n)^2. \blacksquare$$

Proof of lemma 2.18: Clearly

$$F_n \Sigma \overline{F_n} = \left(\frac{1}{n}\sum_{k,s=0}^{n-1} \sigma_{k,s}\exp(i2\pi(sl-jk)/n)\right). \tag{A.5}$$

Equation (2.13) follows easily using $l = j$ in (A.5) and the fact that $\sigma_{k,s} = \sigma_{|k-s|}$. Equation (2.14) follows upon dividing (A.5) into two sums, $\sum_0^{n-1}$ and $\sum_{-n+1}^{-1}$, changing the summation index in the second to $t' = n+t$, and using the fact that $\sigma_k = \sigma_{-k}$. ■

Proof of lemma 2.19: Immediate from the fact that Σ^c_{2n} is a circulant. ■

Proof of lemma 2.20: By Lemma 2.2 f is continous and since it is positive $f(\omega) > \epsilon$ for some $\epsilon > 0$. Now,

$$\begin{aligned} f(\omega) &= \sum_{|k|<n} c_k \exp(i2\pi\omega k) + \sum_{|k|\geq n} c_k \exp(i2\pi\omega k) \\ &= a_n(\omega) + r_n(\omega) \end{aligned}$$

Notice that $a_n(k/2n) = \lambda_k^c$ is the k^{th} eigenvalue of Σ_{2n}^c, and that a_n and r_n are real-valued functions. On the other hand, since $\sum |c_k| < \infty$ there exists $N = N(\epsilon)$ such that $r_n(\omega) < \epsilon/2$ for $n > N$ and thus

$$a_n(\omega) = f(\omega) - r_n(\omega) > \epsilon - \epsilon/2 = \epsilon/2.$$

which means that all the eigenvalues of Σ_{2n}^c are positive for $n > N$. ■

A.2 Chapter 3

A.2.1 Auxiliary definitions and results

This section is devoted to establishing definitions and auxiliary lemmas which we use in deriving the results in Chapter 3.

In what follows, let Ω be complete and separable. Let $\mathcal{B}$ be the Borel σ-algebra and $\mathcal{M}$ the space of probability measures on $(\Omega, \mathcal{B})$. Let $d(x, y) \equiv ||x - y||$ denote the metric on Ω.

Definition A.1 *(Huber, 1981) For any subset $A \subset \Omega$ define the closed δ-neighborhood of A as $A^\delta = \{x \in \Omega \mid \inf_{y \in A} d(x, y) \leq \delta\}$.*

Definition A.2 *(Huber, 1981) The Prohorov distance between two members $F, G \in \mathcal{M}$ is*

$$d_P(F, G) = \inf \{\epsilon > 0 | F\{A\} \leq G\{A^\epsilon\} + \epsilon \text{ for all } A \in \mathcal{B}\}.$$

Theorem A.4 (Strassen) *(Huber, 1981) Let $\epsilon \geq 0$ and $F, G \in \mathcal{M}$. The following two statements are equivalent:*

1. *$d_P(F, G) \leq \epsilon$*
2. *There exist (dependent) random variables U and V with values in Ω such that $\mathcal{L}(U) = F$ and $\mathcal{L}(V) = G$, and $P\{d(U, V) > \epsilon\} \leq \epsilon$.*

The following results are concerned with establishing properties of conditional probabilities of sequences of random variables.

Lemma A.5 *Let X_n and Y_n be sequences of random variables which, for each n, are defined on the same probability space. Then $Y_n = o_p(1)$ implies $P\{||Y_n|| > \epsilon | X_n\} = o_p(1)$.*

Proof of Lemma A.5: $Y_n = o_p(1)$ implies that for all $\epsilon > 0$, there exists $N = N(\epsilon)$ such that $P\{||Y_n|| > \epsilon\} \leq \epsilon^2$ whenever $n \geq N$. By Markov's inequality, for $n \geq N$,

$$P\{P\{||Y_n|| > \epsilon | X_n\} > \epsilon\} \leq \frac{EP\{||Y_n|| > \epsilon | X_n\}}{\epsilon} = \frac{P\{||Y_n|| > \epsilon\}}{\epsilon} \leq \frac{\epsilon^2}{\epsilon} = \epsilon. \blacksquare$$

Lemma A.6 *Let U_n, V_n, and W_n be sequences of random variables which, for each n, are defined on the same probability space. If, for each $\epsilon > 0$, $P\{||U_n - V_n|| > \epsilon | W_n\} = o_p(1)$, then $d_P(\mathcal{L}(U_n|W_n), \mathcal{L}(V_n|W_n)) = o_p(1)$.*

Proof of lemma A.6: Let $X_n = P\{\|U_n - V_n\| > \epsilon | W_n\}$ and $Y_n = d_P(\mathcal{L}(U_n|W_n), \mathcal{L}(V_n|W_n))$. Strassen's Theorem implies that $Y_n \leq \epsilon$ whenever $X_n \leq \epsilon$. It follows that $P\{X_n > \epsilon\} \geq P\{Y_n > \epsilon\}$ but $P\{X_n > \epsilon\} \leq \epsilon$ for n large enough and the lemma follows. ■

Lemma A.7 *If D_n is a positive random variable and $E(D_n|Z_n) = o_p(r_n)$, then $D_n = o_p(r_n)$.*

Proof of Lemma A.7: Let $X_n = E(D_n|Z_n)$. By definition $X_n = o_p(r_n)$ if, and only if, for all $\delta, \epsilon > 0$ there exists $N = N(\delta, \epsilon)$ such that for $n \geq N$, the set $A_n = \{X_n \geq \epsilon\delta r_n\}$ has probability $P\{A_n\} \leq \delta$. For $n \geq N$, we have on A_n^c

$$P\{D_n \geq \epsilon r_n | Z_n\} \leq \frac{E\{D_n|Z_n\}}{\epsilon r_n} \leq \frac{\delta\epsilon r_n}{\epsilon r_n} = \delta.$$

Hence,

$$\begin{aligned} P\{D_n \geq \epsilon r_n\} &= \int P\{D_n \geq \epsilon r_n | Z_n\} dP \\ &\leq \delta + \delta = 2\delta \blacksquare \end{aligned}$$

A.2.2 Proofs for Chapter 3

Proof of Lemma 3.1: By Lemma 2.11

$$\begin{aligned} E\hat{\theta}_x &= \int \theta(\omega) K_n * f(\omega) d\omega = \int \theta(\omega)\{f(\omega) + M_f\, O(\log n/n)\} d\omega \\ &= \int \theta(\omega) f(\omega) d\omega + M_f\, O(\log n/n) \blacksquare \end{aligned}$$

Proof of Lemma 3.2: By Lemma 2.14

$$\begin{aligned} A_n &= \operatorname{cov}(\hat{\theta}_x, \hat{\theta}_y) \\ &= \int\int \theta(\omega)\phi(\omega') cov(I_n^x(\omega), I_{n,\omega'}^y) d\omega d\omega' \\ &= \frac{1}{n^2}\int\int \theta(\omega)\phi(\omega') \left| \int F(x)\overline{G(x)} D_n(x-\omega) D_n(x-\omega') dx \right|^2 d\omega d\omega' \\ &\quad + \frac{1}{n^2}\int\int \theta(\omega)\phi(\omega') \left| \int F(x)\overline{G(x)} D_n(x-\omega) D_n(x+\omega') dx \right|^2 d\omega d\omega' \\ &= B_n + C_n \end{aligned}$$

Expanding the squared term in B_n above

$$\begin{aligned} B_n &= \frac{1}{n^2}\int_\omega \theta(\omega) \int_x \int_y F(x)\overline{G(x)} G(y)\overline{F(y)} D_n(x-\omega) D_n(y-\omega) \\ &\quad \times \int_{\omega'} \phi(\omega') D_n(x-\omega') D_n(y-\omega') d\omega' dx dy d\omega. \end{aligned}$$

By Lemma 2.10 we have that, uniformly in ϕ and θ,

$$\int \phi(\omega') D_n(x-\omega') D_n(y-\omega') d\omega' = \phi(x) D_n(x-y) + M_\phi O(\log n),$$

and analogously

$$\int \theta(\omega) D_n(x-\omega) D_n(y-\omega) d\omega = \theta(x) D_n(x-y) + M_\theta O(\log n).$$

It then follows that, uniformly in F, G, θ, and ϕ,

$$\begin{aligned}
B_n &= \frac{1}{n^2}\int\int F(x)\overline{G(x)F(y)}G(y)\left(\theta(x)D_n(x-y)+M_\theta O(\log n)\right)\times \\
&\quad \times\left(\phi(x)D_n(x-y)+M_\phi O(\log n)\right)dxdy \\
&= \frac{1}{n^2}\int\int F(x)\overline{G(x)F(y)}G(y)\left(\phi(x)\theta(x)nK_n(x-y)\right. \\
&\quad \left.+\,O(\log n)(M_\phi\theta(x)+M_\theta\phi(x))D_n(x-y)+M_\theta M_\phi O(\log^2 n)\right)dydx \\
&= \frac{1}{n}\int F(x)\overline{G(x)}\theta(x)\phi(x)\int\overline{F(y)}G(y)K_n(x-y)dydx \\
&\quad +O(\log n/n^2)\int F(x)\overline{G(x)}\left(M_\phi\theta(x)+M_\theta\phi(x)\right)\int\overline{F(y)}G(y)D_n(x-y)dydx \\
&\quad +O(\log^2 n/n^2)M_\theta M_\phi\left|\int F(x)\overline{G(x)}dx\right|^2 \\
&= \frac{1}{n}\int f(x)g(x)\theta(x)\phi(x)dx+O(\log n/n^2)M_{F\overline{G}}\sup|FG\theta\phi| \\
&\quad +O(\log^2 n/n^2)\sup|FG|^2\left(2\max\{M_\theta\sup|\phi|,M_\phi\sup|\theta|\}+M_\theta M_\phi\right)
\end{aligned}$$

Since θ and ϕ are in Lip(1) and are fixed functions, we shall, for notational convenience, drop uniformity in θ and ϕ and so

$$B_n=\int f(x)g(x)\theta(x)\phi(x)dx+M_{F\overline{G}}\sup|FG|O(\log n/n^2)+\sup|fg|O(\log^2 n/n^2),$$

uniformly in F and G. The result now follows immediately upon noticing that the bound on C_n is the same. ■

Proof of Theorem 3.3: By Lemma 3.1 we have that

$$\begin{aligned}
E|U_n-V_n|^2 &= nE(\hat\theta_x-\theta_f-(\hat\theta_y-\theta_g))^2 \\
&= n\mathrm{Var}(\hat\theta_x-\hat\theta_y)+n\left\{E(\hat\theta_x-\theta_f)+E(\hat\theta_y-\theta_g)\right\}^2 \\
&= n\left(\mathrm{Var}(\hat\theta_x)+\mathrm{Var}(\hat\theta_y)-2\mathrm{cov}(\hat\theta_x,\hat\theta_y)\right) \\
&\quad +(M_g+M_f)^2O(\log^2 n/n)
\end{aligned}$$

so that the theorem follows readily from Lemma 3.2. ■

Proof of Theorem 3.4: The regularity conditions on $\hat f_n$ and Theorem 3.3 imply $E\{|\tilde U_n-\tilde V_n|^2|X_0^{n-1}\}=o_p(1)$ which, by Lemma A.7, in turn implies $|\tilde U_n-\tilde V_n|=o_p(1)$. ■

Proof of Corollary 3.5: (1) follows from the fact that $\tilde U_n-\tilde V_n=o_p(1)$ implies (Lemma A.5) $P\{|\tilde U_n-\tilde V_n|>\epsilon|X_0^{n-1}\}=o_p(1)$ which in turn implies (Lemma A.6) $d_P(\mathcal{L}(\tilde U_n|X_0^{n-1}),\mathcal{L}(\tilde V_n|X_0^{n-1}))=o_p(1)$ but $\mathcal{L}(\tilde U_n|X_0^{n-1})=\mathcal{L}(\tilde U_n)$. (2) follows from the triangle inequality and the fact that $\mathcal{L}(U_n)\to\mathcal{L}(U)$. ■

Proof of Lemma 3.6: Hasminskii's and Ibragimov's (1986) result (3.3), can be used to establish

$$\mathcal{L}(a'\sqrt{n}(\theta_x-\theta_f))\to N(0,a'\Sigma a)$$

for any vector a so that the result follows by taking limits of characteristic functions. ■

Proof of Theorem 3.8: Using a Taylor expansion we can see that

$$\begin{aligned}
\tilde U_n^* &= \sqrt{n}\sum_{j=1}^{p}\frac{\partial h(\theta_f)}{\partial\theta_j}(\hat\theta_{x_j}-\theta_{f_j})+A_n \\
&= \tilde U_n^{**}+A_n
\end{aligned}$$

where

$$A_n \leq \sqrt{n}\|\hat{\theta}_{\hat{s}} - \theta_f\| \left\| \frac{\partial h(\theta_f)}{\partial \theta} - \frac{\partial h(\hat{\theta}^*_{\hat{s}})}{\partial \theta} \right\| = \sqrt{n}\|\hat{\theta}_{\hat{s}} - \theta_f\| A'_n$$

and $\theta^*_{\hat{s}}$ is between $\theta_{\hat{s}}$ and θ_f.

Similarly,

$$\begin{aligned} \tilde{V}^*_n &= \sqrt{n} \sum_{j=1}^{p} \frac{\partial h(\theta_f)}{\partial \theta_j} (\hat{\theta}_{\hat{g}j} - \theta_{f_n j}) + B_n + C_n \\ &= \tilde{V}^{**}_n + B_n + C_n \end{aligned}$$

where

$$B_n \leq \sqrt{n}\|\hat{\theta}_{\hat{g}} - \theta_{f_n}\| \left\| \frac{\partial h(\theta_{f_n})}{\partial \theta} - \frac{\partial h(\hat{\theta}^*_{\hat{g}})}{\partial \theta} \right\| = \sqrt{n}\|\hat{\theta}_{\hat{g}} - \theta_{f_n}\| B'_n,$$

$$C_n \leq \sqrt{n}\|\hat{\theta}_{\hat{g}} - \theta_{\hat{f}_n}\| \left\| \frac{\partial h(\theta_{f_n})}{\partial \theta} - \frac{\partial h(\theta_f)}{\partial \theta} \right\| = \sqrt{n}\|\hat{\theta}_{\hat{g}} - \theta_{\hat{f}_n}\| C'_n,$$

and $\theta^*_{\hat{g}}$ is between $\theta_{\hat{g}}$ and $\theta_{\hat{f}_n}$.

Now, $\tilde{U}^{**}_n - \tilde{V}^{**}_n = o_p(1)$ by Theorem 3.4 and by Theorem 3.7 $\sqrt{n}(\hat{\theta}_{\hat{s}} - \theta_f)$ has a limiting distribution and hence $\sqrt{n}\|\hat{\theta}_{\hat{s}} - \theta_f\| = O_p(1)$. Theorem 3.4 implies that $\sqrt{n}(\hat{\theta}_{\hat{g}} - \theta_{\hat{f}_n})$ has the same limiting distribution as $\sqrt{n}(\hat{\theta}_{\hat{s}} - \theta_f)$ and so $\sqrt{n}\|\hat{\theta}_{\hat{g}} - \theta_{\hat{f}_n}\| = O_p(1)$.

It remains to show that A'_n, B'_n, and C'_n are each $o_p(1)$. $A'_n = o_p(1)$ follows from $\|\hat{\theta}^*_{\hat{s}} - \theta_f\| \leq \|\hat{\theta}_{\hat{s}} - \theta_f\| = O_p(n^{-1/2})$ and the fact that h has continous derivatives in N_{θ_f}. On the other hand, if $d(f, \hat{f}_n) = o_p(1)$, then, for n large, $\theta_{\hat{f}_n} \in N_{\theta_f}$ with high probability, and since h has continous derivatives in N_{θ_f}, this implies that $B'_n = o_p(1)$ and $C'_n = o_p(1)$. ∎

Proof of Corollary 3.9: Theorem 3.8 establishes $\tilde{U}^*_n - \tilde{V}^*_n = O_p(1)$. The rest of the proof is identical to that of Corollary 3.5. ∎

A.3 Chapter 4

A.3.1 Distances between spectra

Definition A.3 *Let f and g be two complex-valued, measurable functions defined on $[0, 1]$. Define the metric $d_2(f, g)^2 = \int_0^1 |f(\omega) - g(\omega)|^2 d\omega$.*

Lemma A.8 *If f and g are complex-valued, measurable functions on $[0, 1]$, then*

$$d_2(f, g) \leq \sup_{\omega} \max(|f(\omega)|, |g(\omega)|) d_2(\log f, \log g).$$

The proof is a consequence of the following lemma:

Lemma A.9 $|1 - e^z| \leq |z| \max(|e^z|, 1)$ *for complex z.*

Proof: Let $z = x + iy$. Then

$$\begin{aligned} |e^z - 1|^2 &= e^{2z} - 2e^z \cos(y) + 1 && \text{(A.6)} \\ &= (e^z - 1)^2 + 2e^z(1 - \cos(y)) && \text{(A.7)} \\ &\leq (e^z - 1)^2 + e^z y^2 && \text{(A.8)} \end{aligned}$$

since $1 - \cos(y) \leq y^2/2$. By Taylor's theorem, there exists $0 \leq \theta \leq 1$ such that $(e^x - 1)^2 = x^2 e^{\theta 2x} \leq x^2 \max(e^{2x}, 1)$. Hence $|e^z - 1|^2 \leq (x^2 + y^2)\max(e^{2x}, 1)$. The result follows upon taking square root and using the fact that $|e^z| = e^x$. ■

We can now prove lemma A.8: Apply Lemma A.9 with $z = \log g - \log f$:

$$|f - g| = |f||1 - g/f| \leq |\log f - \log g| \max(|f|, |g|). \tag{A.9}$$

The result now follows upon integrating A.9. ■

A.3.2 Proofs for Chapter 4

Proof of lemma 4.1: Lemmas 2.11 and 2.13 show that

$$V_{kn} = (1 + \rho_k \theta_k)(U_{1k}^2 + U_{2k}^2)/2$$

where U_{ik} are iid $N(0,1)$ variates, θ_k is a random variable such that $|\theta_k| \leq 1$, and ρ_k is the correlation between the real and imaginary parts of the DFT. Thus we have that

$$\epsilon_{kn} = \log(U_{1k}^2 + U_{2k}^2)/2 + \log(1 + \rho_k \theta_k) + \gamma.$$

Lemma 2.12 establishes $\rho_k = M_f O(\log n/n)$. Since $(U_{1k}^2 + U_{2k}^2)/2$ is an exponential random variable, the expected value of its log is just $-\gamma$ so that

$$E\epsilon_{kn} = \rho_k E\theta_k - \rho_k^2 E\theta_k^2/2 + \rho_k^4 E\theta_k^3/3 + O(\rho_k^4).$$

The first assertion then follows from the fact that $E(\theta_k^m) = 0$ for m odd and the second follows trivially.

Now, the variance of

$$\epsilon_{kn} = \log(1 + \rho_k \theta_k) + \log(U_{1k}^2 + U_{2k}^2)/2 + \gamma$$

is equal to the sum of the variances of the first two terms since lemma 2.13 shows they are independent. The second term is just the log of an exponential so its variance is known to be $\pi^2/6$. Moreover, the variance of the first term is bounded by $(\log(1 - \rho_k))^2 = O(\rho_k^2)$ so that the third assertion follows trivially.

The fourth assertion is merely a restatement of lemma 2.17. ■

Proof of lemma 4.2: Notice that

$$\sup_\omega |W_{\lambda,m,n}^{(k)}(\omega)| \leq \frac{2 - \delta_k}{n} \sup_{x \geq 0} \frac{(2\pi x)^k}{1 + \lambda(2\pi x)^{2m}} + \frac{2}{n} \int_0^\infty \frac{(2\pi x)^k dx}{1 + \lambda(2\pi x)^{2m}}$$

The result now follows upon changing variable of integration to $y = \lambda^{1/2m} 2\pi x$ in the integral above and verifying that

$$\sup_{x \geq 0} \frac{(2\pi x)^k}{1 + \lambda(2\pi x)^{2m}} = (1 - k/2m) \left(\frac{k}{\lambda(2m - k)} \right)^{k/2m} = O(\lambda^{-k/2m}). ■$$

Proof of Lemma 4.3: By Lemma 4.2 and since $f \in \text{Lip}(1)$ we have that, for n large,

$$\begin{aligned} \tilde{h}_n(\omega) &= \sum_{k=-n+1}^{n} \{\log f(k/2n) + O(\log n/n)\} W(\omega - k/2n) \\ &= \sum_{k=-n+1}^{n} \log f(k/2n) W(\omega - k/2n) + (1 + \lambda^{-1/2m}) O(\log n/n) \\ &= B_n(\omega) + (1 + \lambda^{-1/2m}) O(\log n/n), \end{aligned}$$

so that it suffices to show $\sup_\omega |B_n(\omega)| \le \sum |b_\nu|$. Indeed,

$$\begin{aligned} B_n(\omega) &= \frac{1}{2n} \sum_{\nu=-\infty}^{\infty} b_\nu \sum_{s=-n}^{n} \frac{\exp(i2\pi s\omega)}{1+\lambda(2\pi s)^{2m}} \sum_{k=-n+1}^{n} \exp(i2\pi k(\nu - s)/2n) \\ &= \sum_{\nu=-n}^{n} \frac{b_\nu \exp(i2\pi\nu\omega)}{1+\lambda(2\pi\nu)^{2m}} \qquad (A.10) \\ &\le \sum_{\nu=-\infty}^{\infty} |b_\nu|. \blacksquare \end{aligned}$$

Proof of Lemma 4.4: A similar derivation to that used in deriving (A.10) in the proof of Lemma 4.3, shows that

$$\tilde{h}_n(\omega) = \sum_{\nu=-n}^{n} \frac{b_{\nu n} \exp(i2\pi\nu\omega)}{1+\lambda(2\pi\nu)^{2m}}$$

where $\{b_{\nu n}\}$ are the Fourier coefficients of $\log(K_{2n} * f)$. This implies that

$$\tilde{h}_n(\omega) = 2nW_{\lambda,m,n} * \log(K_{2n} * f)(\omega)$$

and hence by Lemmas 2.5, 2.6, and 4.2, we have that

$$\begin{aligned} \sup_\omega |\tilde{h}_n'(\omega)| &\le 2n \sup_\omega \left| \frac{K_{2n} * f'(\omega)}{K_{2n} * f(\omega)} \right| \|W_{\lambda,m,n}\|_1 \\ &\le \sup_\omega |f'(\omega)| \sup_\omega \frac{1}{f(\omega)} (1+\lambda^{-1/2m}) O(1). \end{aligned}$$

uniformly in λ. But f' is bounded since $f \in \mathrm{Lip}(1)$ is differentiable. Furthermore, since $f > 0$, and is continuous, $1/f$ is bounded. ■

Proof of Lemma 4.5: The r.h.s. in 4.6 follows from the fact that (cf., e.g., the proof of Lemma 4.3)

$$\tilde{h}_n(\omega) = \sum_{\nu=-n}^{n} \frac{b_\nu \exp(i2\pi\nu\omega)}{1+\lambda(2\pi\nu)^{2m}} + (1+\lambda^{-1/2m}) O(\log n/n),$$

Parseval's theorem, and the fact that (cf., e.g., Butzer and Nessel, 1971)

$$\sum_{|\nu|>n} |b_\nu|^2 = \|D_{2n+1} * h - h\|_2^2 = O(n^{-2})$$

whenever $h \in \mathrm{Lip}(1)$. Finally, continuity of $A(\lambda)$ follows from a discrete version of Lebesgue's dominated convergence theorem. ■

Proof of Lemma 4.6: From (4.4) and (4.5) we have that

$$\begin{aligned} h_n^*(\omega) &= \frac{1}{2n} \sum_{k=-n+1}^{n} \epsilon_{kn} \sum_{\nu=-n}^{n} \frac{\exp(i2\pi\nu(\omega - k/2n))}{1+\lambda(2\pi\nu)^{2m}} \\ &= \sum_{\nu=-n}^{n} \frac{\exp(i2\pi\nu\omega)}{1+\lambda(2\pi\nu)^{2m}} \left(\frac{1}{2n} \sum_{k=-n+1}^{n} \epsilon_{kn} \exp(-i2\pi\nu k/2n) \right) \\ &= \sum_{\nu=-n}^{n} \frac{\exp(i2\pi\nu\omega) B_{\nu n}}{1+\lambda(2\pi\nu)^{2m}} \end{aligned}$$

Now, by Lemma 4.1

$$\begin{aligned} E|B_{\nu n}|^2 &= \frac{1}{4n^2}\sum_{j,k=-n+1}^{n} E(\epsilon_{kn}\epsilon_{jn})\exp(-i2\pi\nu(k-j)/2n) \\ &\leq O(n^{-1}) + O(\log^2 n/n^2) = O(n^{-1}) \end{aligned}$$

uniformly in ν.

On the other hand, by Parseval's Theorem

$$\begin{aligned} \int E\{h_n^*(\omega)^2\}\,d\omega &= \sum_{\nu=-n}^{n} \frac{EB_{\nu n}^2}{(1+\lambda(2\pi\nu)^{2m})^2} \\ &= O(n^{-1})\sum_{\nu=-n}^{n}\frac{1}{(1+\lambda(2\pi\nu)^{2m})^2} \\ &= O(n^{-1})(1+\lambda^{-1/2m}) \end{aligned}$$

by bounding the sum by an integral and changing the variable of integration as in the proof of Lemma 4.2. ∎

Proof of Theorem 4.7: Take $0 < s < 1$ and divide up the $2n$ integers from $-n+1$ to n into $J \simeq n^{1-s}$ consecutive disjoint sets E_j from $a_1 + 1 = -n+1$ to a_2, a_2+1 to a_3, etc., each set of length n^s or less. Let

$$A_{jn} = \sum_{k\in E_j}\epsilon_{kn},$$

$1 \leq j \leq J$. Then,

$$EA_{jn} = O\left(n^s\frac{\log^2 n}{n^2}\right) = O\left(\frac{\log^2 n}{n^{2-s}}\right),$$

and

$$\mathrm{Var}(A_{jn}) \leq n^s O(1) + n^{2s}O(\log^2 n/n^2) = O(n^s),$$

where O is uniform in j for $1 \leq j \leq J$. Applying Markov's inequality to A_{jn}^2 we have that

$$P\left\{|A_{jn}| > n^{1/2+\delta}\right\} \leq \frac{O(n^s) + O(\log^2 n/n^{2-s})^2}{n^{1+2\delta}} = O(n^{s-1-2\delta})$$

and if $A_n = \max_{1\leq j\leq J}|A_{jn}|$ then

$$\begin{aligned} P\left\{A_n > n^{1/2+\delta}\right\} &\leq \sum_j P\left\{|A_{jn}| > n^{1/2+\delta}\right\} \\ &= \frac{JO(n^s)}{n^{1+2\delta}} = \frac{O(n^{1-s}n^s)}{n^{1+2\delta}} = O(n^{-2\delta}) \to 0. \end{aligned}$$

Thus $A_n = o_p(n^{1/2+\delta})$, for each $\delta > 0$. Now,

$$\begin{aligned} h_n^*(\omega) &= \sum_{j=1}^{J}\sum_{k\in E_j} W_n(\omega - k/2n)\epsilon_{kn} \\ &= \sum_{j=1}^{J}\left\{W_n(\omega - a_j/2n)A_{jn} + \sum_{k\in E_j}(W_n(\omega-k/2n) - W_n(\omega - a_j/2n))\epsilon_{kn}\right\} \end{aligned}$$

The first sum above, S_{1n}, satisfies

$$\begin{aligned} S_{1n} &\leq A_n \sup_\omega \sum_{j=1}^{J} |W_n(\omega - a_j/2n)| \\ &\leq A_n \sup_\omega |W_n(\omega)| J \\ &= o_p(n^{1/2+\delta}) O(n^{r_0-1}) O(n^{1-s}) \\ &= O_p(n^{1/2+\delta+r_0-s}) \end{aligned}$$

and the second sum, S_{2n}, satisfies

$$\begin{aligned} S_{2n} &\leq \sum_{j=1}^{J} \sum_{k \in E_j} |W_n(\omega - k/2n) - W_n(\omega - a_j/2n)||\epsilon_{kn}| \\ &\leq \sum_{j=1}^{J} \frac{n^s}{n} O(n^{r_1-1}) \sum_{k \in E_j} |\epsilon_{kn}| \\ &\leq \frac{n^s}{n} O(n^{r_1-1}) \sum_j \sum_{k \in E_j} |\epsilon_{kn}| \\ &= nO_p(n^{s+r_1-2}) \\ &= O_p(n^{s+r_1-1}) \end{aligned}$$

since $\sum_k |\epsilon_{kn}| = O_p(n)$ by Markov's inequality. This proves the first assertion of the lemma. An analogous argument proves the second. ■

Proof of Lemma 4.8: By the triangle inequality

$$\begin{aligned} d_2(h, \hat{h}_n) &= d_2(h, \tilde{h}_n + h_n^*) \\ &\leq d_2(h, \tilde{h}_n) + d_2(0, h_n^*) \end{aligned}$$

and the result follows immediately from Lemmas 4.5 and 4.6 since $Ed_2(h_n^*, 0) \leq \sqrt{Ed_2^2(h_n^*, 0)}$. ■

Proof of Lemma 4.9: Apply Lemma A.8 to f and $\hat{f}_n$. The O_p term follows since $\hat{f}_n$ is bounded in probability (Lemma 4.7). The expression in parentheses follows from Lemma 4.8. ■

A.4 Chapter 7

Proof of Lemma 7.1: Since $V = \mathrm{mod}(U + a, 1) \sim U(0,1)$ for all a whenever $U \sim U(0,1)$, we have that

$$Y_0^{n-1} = \overline{F_n} \Lambda_n^{1/2} \Delta_0^{n-1}$$

has the same distribution as

$$\overline{F_n} \Lambda_n^{1/2} D_n \Gamma_0^{n-1}$$

where D_n is a diagonal matrix with entries $\exp(i2\pi U_0), \ldots, \exp(i2\pi U_{n-1})$, $U's$ iid $U(0,1)$ so let

$$\begin{aligned} Y_0^{n-1} &= \overline{F_n} \Lambda_n^{1/2} D_n \Gamma_0^{n-1} \\ &= \overline{F_n} D_n \Lambda_n^{1/2} \Gamma_0^{n-1} \\ &= \overline{F_n} D_n F_n X_0^{n-1} \end{aligned}$$

so that the covariance matrix of Y_0^{n-1} given D_n is given by

$$E\left\{Y_0^{n-1}(Y_0^{n-1})^T | D_n\right\} = \overline{F_n} D_n F_n \Sigma_n \overline{F_n D_n F_n}. \tag{A.11}$$

On the other hand, for any matrix $A_n = (a_{jk})$, $D_n A_n \overline{D_n} = (a_{jk} \exp(i2\pi(U_j - U_k)))$, and since $E \exp(i2\pi(U_j - U_k)) = 1$ only if $j = k$ and 0 otherwise, $E D_n A_n \overline{D_n} = \mathrm{diag}(A_n)$, i.e., only the diagonal survives premultiplying by D_n, postmultiplying by $\overline{D_n}$, and taking expectation. Thus, taking expectation again in (A.11) yields

$$\begin{aligned} E\left\{Y_0^{n-1}(Y_0^{n-1})^T\right\} &= \overline{F_n}\,\mathrm{diag}(F_n \Sigma_n \overline{F_n}) F_n \\ &= \Sigma_n^c \end{aligned}$$

where Σ_n^c is as in Example 1 of circulants in Chapter 2. ■

References

Abramowitz and Stegun (1954). *Handbook of Mathematical Functions.* National Bureau of Standards, Washington D.C.

Butzer, P.L and Nessel, R.J. (1971). *Fourier Analysis and Approximation.* Academic Press, New York.

Carlstein, E. (1986). "The use of subseries values for estimating the variance of a general statistic from a stationary sequence". *Ann. Statist.* **14**: 1171-1194.

Efron, B. (1979). "Bootstrap methods: another look at the jacknife". *Ann. Statist.* **7**: 7-126.

Freedman, D.A. and Peters, S.F. (1984). "Bootstrapping an economic model: some empirical results". *Journal of Business and Economic Statistics* **2**: 150-158.

Hasminskii, R.Z. and Ibragimov, I.A. (1986). "Asymptotically efficient nonparametric estimation of functionals of a spectral density function." *Probability Theory and Related Fields* **73**: 447-461.

Huber, P. (1981). *Robust Statistics.* Wiley, New York.

Hurvich, C.M., and Zeger, S. (1986). "Frequency domain bootstrap methods for time series". Unpublished manuscript. Department of Statistics and Operations Research, New York University, and Department of Biostatistics, Johns Hopkins University.

Johnson, N.L. and Kotz, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions.* Wiley, New York.

Kendall, M. & Stuart, A. (1976). *The Advanced Theory of Statistics, vol 3, third ed.* Charles Griffin and Co. Ltd., London

Ramos, E (1984). "A bootstrap for time series". Qualifying Paper. Department of Statistics, Harvard University.

Rao, C.R. (1973). *Linear Statistical Inference and its Applications, 2nd ed..* Wiley, New York.

Stine, R. (1985). "A frequency domain bootstrap for time series". Unpublished manuscript. Department of Statistics, Wharton School, University of Pennsylvania.

Wahba, G. (1980). "Automatic smoothing of the log periodogram". *JASA* **75**: pp. 122-132.

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION / AVAILABILITY OF REPORT unlimited
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) TR NO. ONR-C-2	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Dept. of Statistics Harvard University 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) Dept. of Statistics SC713 Harvard University Cambridge, MA 02138	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING / SPONSORING ORGANIZATION ONR 8b. OFFICE SYMBOL (If applicable) Code 1111	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-86-K-0246
8c. ADDRESS (City, State, and ZIP Code) Office of Naval Research Arlington, VA 22217-5000	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO

11. TITLE (Include Security Classification)

Resampling Methods for Time Series

12. PERSONAL AUTHOR(S)
Ernesto Ramos

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Technical Report	FROM _____ TO _____	August 27, 1988	57

16. SUPPLEMENTARY NOTATION

17 COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	TIME SERIES, BOOTSTRAP, RESAMPLING, SPECTRAL ESTIMATION

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

See reverse side.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Herman Chernoff	22b. TELEPHONE (Include Area Code) 617-495-5462 / 22c. OFFICE SYMBOL

Abstract

Suppose we observe a realization of size n of a Gaussian stationary sequence and we estimate θ_f, a functional of the spectral density, by $\hat{\theta}_x$. For example, θ_f and $\hat{\theta}_x$ might be the true and the empirical rate at which the process crosses a certain level. It is natural to ask: (i) can the sampling distribution of $\hat{\theta}_x$ be consistently estimated? and (ii), can a better estimator of θ_f than $\hat{\theta}_x$ be constructed on the basis of $\hat{\theta}_x$?

In this paper we describe approaches which allow answering both questions affirmatively in certain situations. The approach is based on *resampling* the data, i.e., using the original data to specify a distribution from which new samples are drawn. In the first problem the data induce an estimate of the spectral density $\hat{f}_n$ and the sampling distribution of $\hat{\theta}_x$ under $\hat{f}_n$ is used to estimate that under f. We establish results for estimators of well-behaved functions of linear functionals of the spectral density. In the second problem we pretend the data are circular so that the *periodogram* of the data is a sufficient statistic and thus the conditional expectation of $\hat{\theta}_x$ is a better estimator. If the data are not really circular, this mechanism introduces bias but the reduction in variance may be substantial enough to reduce the mean square error.